PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N

A2

(11) International Publication Number: WO 00/43494

(43) International Publication Date: 27 July 2000 (27.07.00)

(21) International Application Number: PCT/US00/01849

(22) International Filing Date: 27 January 2000 (27.01.00)

(30) Priority Data:
60/117,285 26 January 1999 (26.01.99) US

(71) Applicant *(for all designated States except US)*: THE SCRIPPS RESEARCH INSTITUTE [US/US]; Office of Patent Counsel, 10550 North Torrey Pines Road, Mail TPC–8, La Jolla, CA 92037 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: FAUQUET, Claude [FR/US]; 308 South Spoede Road, St. Louis, MO 63141 (US). CHATTERJI, Anju [IN/US]; 613C Westgate Avenue, University City, MO 63130 (US).

(74) Agents: ESMOND, Robert, W. et al.; Sterne, Kessler, Goldstein & Fox P.L.L.C., Suite 600, 1100 New York Avenue, N.W., Washington, DC 20005–3934 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**
*Without international search report and to be republished upon receipt of that report.*

(54) Title: CONTROL OF VIRUS INFECTION USING REPLICATION ASSOCIATED PROTEINS, COMPOSITIONS AND METHODS OF USE

(57) Abstract

The invention is generally directed to methods and compositions for controlling infection of plants by geminiviruses using replication associated proteins. In particular, the invention relates to iteron and Rep DNA and polypeptide sequences as well as transgenic plants expressing such sequences.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AL** | Albania | **ES** | Spain | **LS** | Lesotho | **SI** | Slovenia |
| **AM** | Armenia | **FI** | Finland | **LT** | Lithuania | **SK** | Slovakia |
| **AT** | Austria | **FR** | France | **LU** | Luxembourg | **SN** | Senegal |
| **AU** | Australia | **GA** | Gabon | **LV** | Latvia | **SZ** | Swaziland |
| **AZ** | Azerbaijan | **GB** | United Kingdom | **MC** | Monaco | **TD** | Chad |
| **BA** | Bosnia and Herzegovina | **GE** | Georgia | **MD** | Republic of Moldova | **TG** | Togo |
| **BB** | Barbados | **GH** | Ghana | **MG** | Madagascar | **TJ** | Tajikistan |
| **BE** | Belgium | **GN** | Guinea | **MK** | The former Yugoslav Republic of Macedonia | **TM** | Turkmenistan |
| **BF** | Burkina Faso | **GR** | Greece | | | **TR** | Turkey |
| **BG** | Bulgaria | **HU** | Hungary | **ML** | Mali | **TT** | Trinidad and Tobago |
| **BJ** | Benin | **IE** | Ireland | **MN** | Mongolia | **UA** | Ukraine |
| **BR** | Brazil | **IL** | Israel | **MR** | Mauritania | **UG** | Uganda |
| **BY** | Belarus | **IS** | Iceland | **MW** | Malawi | **US** | United States of America |
| **CA** | Canada | **IT** | Italy | **MX** | Mexico | **UZ** | Uzbekistan |
| **CF** | Central African Republic | **JP** | Japan | **NE** | Niger | **VN** | Viet Nam |
| **CG** | Congo | **KE** | Kenya | **NL** | Netherlands | **YU** | Yugoslavia |
| **CH** | Switzerland | **KG** | Kyrgyzstan | **NO** | Norway | **ZW** | Zimbabwe |
| **CI** | Côte d'Ivoire | **KP** | Democratic People's Republic of Korea | **NZ** | New Zealand | | |
| **CM** | Cameroon | | | **PL** | Poland | | |
| **CN** | China | **KR** | Republic of Korea | **PT** | Portugal | | |
| **CU** | Cuba | **KZ** | Kazakstan | **RO** | Romania | | |
| **CZ** | Czech Republic | **LC** | Saint Lucia | **RU** | Russian Federation | | |
| **DE** | Germany | **LI** | Liechtenstein | **SD** | Sudan | | |
| **DK** | Denmark | **LK** | Sri Lanka | **SE** | Sweden | | |
| **EE** | Estonia | **LR** | Liberia | **SG** | Singapore | | |

# Control of Virus Infection Using Replication Associated Proteins, Compositions and Methods of Use

## *Background of the Invention*

### *Field of the Invention*

The invention relates to methods and compositions for controlling virus infection using replication associated proteins.

### *Related Art*

Viruses can infect both animals and plants. Plant viruses damage plants following infection, and are the cause of substantial agricultural losses. Geminiviruses in particular are extremely devastating plant viruses found all over the world. Natural resistance genes are rare and usually not in the plant species needed and modern biotechnology (including genetic engineering) has not yet provided extremely effective measures to control the geminiviruses, as well as other viruses whose replication involves binding of a replication associated protein (Rep) to an iteron.

#### *Geminiviruses*

Geminiviruses belong to a family of plant viruses that cause economically important diseases in a wide range of cereal, vegetables and fiber crops (Brown, J. K., *FAO Plant Prot. Bull. 42*:3-32 (1994)). They are characterized by twinned icosahedral particle morphology and covalently closed circular, single stranded

DNA as their genome (Stanley, J., *Virol. 2*:139-150 (1991); Lazarowitz, S., *Crit. Rev. Plant. Sci. 11*:29-349 (1992)). Geminiviruses have a single stranded DNA genome that replicates in the infected nuclei by a rolling circle mechanism (Stenger, D.C., *et al., Proc. Natl. Acad. Sci. USA 88*:8029-8033 (1991); Saunders, K., *et al., Nucleic. Acids. Res.19*:2325-2330 (1991)).

The geminiviruses are transmitted by leafhoppers or whitefly vectors and have monopartite or bipartite genomes. Geminiviruses with a bipartite genome have their essential viral functions divided on two DNA components referred to as DNA-A and DNA-B. The DNA-A encodes the replication associated protein (Rep), the replication enhancer (REn), the transcriptional activator protein (TrAP), and the coat protein (CP), while the movement functions are located on DNA-B. Genetic studies have shown that both the movement protein (MP) and the nuclear shuttle protein (NSP) encoded by DNA-B are necessary for systemic infection (Brough, C.L., *et al., J. Gen. Virol. 69*:503-514 (1988); Etessami, E., *et al., J. Gen. Virol. 72*:1005-1012 (1991)). In DNA-A and DNA-B the open reading frames (ORFs) are arranged in two divergent clusters separated by an intergenic region (IR) of about 200 nucleotides. The IR contains sequences that are conserved between the two DNA components and are referred to as the common region (CR). The CR contains the origin of replication (*ori*) sequences that are crucial to initiate replication and consists of a conserved hairpin structure and a binding site for the Rep protein located upstream of the hairpin (Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994); Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994)). The *ori* for Squash leaf curl virus (SqCLV) has been mapped to a 90 nucleotide DNA segment within the CR (Lazarowitz, S.G., *et al., Plant Cell 4*:799-809 (1992)) and within a 100 bp fragment in tomato golden mosaic virus (TGMV, Fontes, E.P.B., *et al., Plant Cell 4*:597-608 (1992)).

Replication of the viral DNA occurs in the nuclei of infected cells via double stranded DNA intermediates (Saunders, K., *et al., Nucleic Acids Res. 19*:2325-2330 (1991); Stenger, D.C., *et al., Proc. Natl. Acad. Sci. USA 88*:8029-8033 (1991)). The IR contains a GC rich inverted repeat which is conserved in all geminiviruses and has the potential to form a stem-loop structure.

These inverted repeats flank an AT rich sequence of 11-16 bases that contains the conserved nonamer motif, TAATATTAC. (SEQ ID NO: 187)

*Replication Associated Protein (Rep)*

Of the different gene products encoded by the geminivirus, only AC1 or the replication-associated protein (Rep) is essential for viral DNA replication. This protein is encoded by the ORF AC1 and initiates rolling circle replication by a site specific cleavage within the loop of the conserved structure (Laufs, J., *et al., FEBS Lett. 377*:258-262 (1995a); Laufs, J., *et al., Proc. Natl. Acad. Sci. USA 92*:3879-3883 (1995b)). Rep is a multifunctional protein and is involved in both viral replication and transcriptional regulation (Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994a); Eagle, P.A., *et al., Plant Cell 6*:1157-1170 (1994); Eagle, P.A., and Hanley-Bowdoin L., *J. Virol. 71*:6947-6955 (1997)). The amino terminal region of the Rep protein has been shown to be important for DNA recognition and binding (Choi, I.R., and Stenger D.C., *Virology 206*:904-912 (1995); Choi, I.R., and Stenger D.C., *Virology 226*:22-126 (1996); Jupin, I., *et al., FEBS Lett. 262*:116-120 (1995)) and in cleavage and ligation of the viral origin of replication. The carboxy terminal region of the Rep protein has a nucleoside triphosphate binding domain (Laufs, J., *et al., Proc. Natl. Acad. Sci. USA 92*:3879-3883 (1995b)).

Rep protein possesses a nicking-closing activity and initiates rolling circle replication by a site specific cleavage within the loop of the conserved nonamer sequence, TAATATTAC (Laufs, J.S., *et al., FEBS Lett. 377*:258-262 (1995); Laufs, J., *et al., Proc. Natl. Acad. Sci. USA 92*:3879-3883 (1995); Heyraud-Nitschke, F., *et al., Nucleic Acid Res. 23*:910-916 (1995); Orozco, B.M., and Hanley-Bowdoin, L., *J. Virol 270*:148-158 (1996)). The Rep protein binding site is located between the TATA box and the transcription start site for the Rep gene and acts as the origin recognition sequence and as a negatively regulatory element for Rep gene transcription (Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994); Eagle, P.A., *et al., Plant Cell 61*:1157-1170 (1994); Eagle, P.A., and Hanley-Bowdoin, L., *J. Virol. 71*:6947-6955 (1997))

Recently, discrete functional domains that are responsible for protein binding, cleavage and oligomerization have been identified in the N-terminus of Rep protein of tomato golden mosaic virus, (TGMV, Orozco, B.M., *et al., J. Biol. Chem. 272*:9840-9846 (1997); Orozco, B. M., *et al., Virology 242*:346-356 (1998); Gladfelter, H.J., *et al., Virology 239*:186-197 (1997)).

The high affinity binding sites for the Rep protein of TGMV, bean golden mosaic virus, BGMV (Eagle, P.A., *et al., Plant Cell 6*:1157-1170 (1994); Fontes, E.P.B., *et al., Plant Cell 4*:597-608 (1992); Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994a); Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994b)) and tomato leaf curl virus, TLCV (Behjatania, S.A., *et al., Nucleic Acids Res. 26*:925-931(1998)) have been mapped in the viral origin close to the TATA box and the conserved hairpin structure.

Rep interacts with at least two different DNA elements in the geminivirus origin of replication, a conserved nonanucleotide sequence containing a specific nick site for the enzyme (Heyraud-Nitschke, F., *et al., Nucleic Acids Res. 23*:910-916 (1995); Laufs, J., *et al., FEBS Lett. 377*:258-262 (1995a) Laufs, J., *et al., Proc. Natl. Acad. Sci. USA 92*:3879-3883 (1995b)) and a directly repeated sequence motif located between the TATA box in the promoter of the AC1 gene and the transcription start site (Fontes, E.P.B., *et al., Plant Cell 4*:597-608 (1992); Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994a); Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994b); Choi, I.R., and Stenger D.C., *Virology 226*:22-126 (1996)).

The Rep proteins encoded by different geminiviruses show specificity for the replication of their cognate genomes (Lazarowitz, S.G., *et al., Plant Cell 4*:799-809 (1992); Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994b); Jupin, I., *et al., FEBS Lett. 262*:116-120 (1995); Choi, I.R., and Stenger D.C., *Virology 226*:22-126 (1996)). This specificity of origin recognition is determined in part by the high affinity binding site of the Rep (Choi, I.R., and Stenger D.C., *Virology 206*:904-912 (1995); Choi, I.R., and Stenger D.C., *Virology 226*:22-126 (1996)) and the N-terminal domain of the Rep protein. In the case of tomato yellow leaf

curl virus (TYLCV), the N-terminal domain comprises the first 116 amino acid residues of the Rep protein (Jupin, I., *et al., FEBS Lett. 262*:116-120 (1995)).

Attempts have been made to modify the viral genes involved in viral replication to obtain a virus-resistant plant. In one instance, there has been a disclosure of the use of a modified gene encoding the AL1 structural protein, i.e. a replicase, and its effect on viral replication. U.S. Patent No. 5,850,023.

*Iterons*

Based on a phylogenetic and structural analysis of the IR from 30 different dicot infecting geminiviruses (Arguello-Astorga, G.R., *et al., Virology 203*:90-100 (1994)) identified a series of sequence elements 6 to 12 nucleotides in length which are repeated 3 to 6 times within the origin of replication. Their research indicated that the nucleotide sequence of the iterated elements (iterons) is generally virus specific and proposed that these iterons may represent specific sites for binding of Rep protein. Further, they showed that the orientation, sequence and arrangement of the iterons was highly conserved between different subgroups of the family geminiviridae. In many viruses, both those that infect plants and animals there is an organization of replication that is similar to the geminiviruses. These other viruses include *Nanoviruses and Ciroviridae.*

The first step in the replication process of geminiviruses involves recognition of the iterons in the common region of the virus by the Rep protein. These iterons mostly occur as direct repeat motifs of 6-12 bp within the common region of the viral genome between the TATA box and the start site for the transcription of the AC1 gene. The iterons have been proposed to serve as the high affinity binding sites of the Rep protein and therefore function as origin recognition sequence. Specific regions on the N-terminus of the AC1 gene involved in DNA binding have been identified for tomato golden mosaic virus TGMV (Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994a); Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994b)) and ACMV (Hong, Y., and Stanley, J., *J.Gen.Virol. 76*:2415-2422 (1995)).

The list of iteron sequences and the corresponding N-terminal sequences of the Rep protein from different viruses reiterates that variability at the iteron levels in geminiviruses is limited (*See* Figure 1A-1C). Even though the viruses may originate in different geographical regions or have diverse host range they share similar iteron sequences. Similarly, the sequence of the amino acid residues that might be involved in origin recognition of binding show limited variability and have certain amino acids in the N-terminal sequence of the Rep protein that give them a common base. For instance, all geminiviruses with the iteron GGTAC (SEQ ID NO. 188) in the origin, have the amino acids FQIN (SEQ ID NO: 148) at [psotopms 7-10 on the N-terminus of the Rep protein common between them, other than the motif 1, FLTY (SEQ ID NO: 189 ), which is conserved throughout the geminiviridae. In other cases, the homology is not so remarkable, nevertheless, the 4th residue downstream of the conserved phe (F), is almost always identical between different viruses having the same iteron sequence.

Recent reports involving the formation of pseudorecombinants between closely related viruses provide evidence in favor of the observations that viruses sharing same iteron sequence can form pseudorecombinants as demonstrated in case of CaLCV and SqLCV (Hill *et al*, *Virology, 250*: 283-292 (1998)) between BDMV and ToMoV (Hou *et al*, *J.Virol, 70*:5430-5436 (1997)) and SiGMV/Ho and SiGMV/CR (Frishmuth *et al, J Gen.Virol, 78:* 2675-2682. (1997)). These data strengthen the view that the iteron and the N-terminal sequences in the Rep are two key components that influence origin recognition and replication of the virus, therefore, a control strategy based upon blocking this fundamental step offers a broad spectrum of resistance effective against all geminiviruses having the same iterons. Further, other viruses such as for example the *Nanovirus* or *Circoviruses* that have similar replication mechanisms may also be inhibited by blocking a similar step in the process.

*Nanoviruses*

*Nanovirus is a* genus of virus that includes plant infecting viruses with a genome consisting of a multiple (at least 6) circular ssDNA molecules each of

approximately 1 kb in size and encapsidated in an icosahedral (non-geminate) virion about 20 nm in diameter. It includes species such as the Subterranean clover virus (SCSV).

The virions are 17 to 20 nm in diameter and exhibit icosahedral symmetry. They are not enveloped. Capsomeres may be evident, producing an angular or hexagonal outline. They have buoyant density of 1.24 to 1.30 g/cm$^3$ in $Cs_2SO_4$, and 1.34 g/cm$^3$ in CsCl. Instability in CsCl has been reported for SCSV. An $S_{20w}$ of 46S has been reported for Banana bunchy top virus (BBTV). Particle morphology is not affected by freezing of tissue before virion extraction.

The nanovirus genome is composed of several species of circular ssDNA ranging in size from 985 to 1111 nts. All of them seem to be structurally similar in being positive sense, transcribed in one direction, and containing a conserved stem-loop structure (and other conserved domains) in the non-coding region. Six to 10 DNA components, each of which appears to be encapsidated in a separate particle, have been isolated from virion preparations of different species and their isolates. The number and types of ssDNA components constituting the integral genome parts have not been determined yet for any of the nanoviruses. The virions have a single capsid protein with a Mr of 19 - 20 x $10^3$.

The genomic information of the nanovirus is distributed over at least 6 molecules of circular ssDNA. Since the the nanovirus DNAs are structurally similar to those of the geminiviruses and at least one of the DNA components of each species codes for a replication-associated protein (Rep), nanovirus DNAs are proposed to be replicated from transcriptionally and replicationally active dsDNA forms via a rolling circle type of replication mechanism. Nicking and joining activity of the BBTV Rep protein has been demonstrated in vitro. Complementary strand synthesis of BBTV genomic ssDNA is attributed to a population of endogenous primers derived from BBTV-DNA 5, which appears to encode a protein that is potentially involved into cell cycle regulation.

All ssDNAs found associated with the assigned species contain a major virion sense ORF and appear to be transcribed unidirectionally. Each coding region is preceded by a promoter sequence with a TATA box and followed by a

poly(A) addition signal. At least one of the genome components codes for a Rep protein (Rep; Mr 32.4 - 33.6 x $10^3$). For some isolates of the four assigned species, two to four Rep components have been described, however, some of the additional Rep components may actually be satellite components. A second virion-sense ORF, completely within the Rep ORF and encoding a putative 5 x $10^3$ protein of unknown function, was also identified for the BBTV Rep component (DNA-1), which is present in all BBTV isolates studied. Available information suggests that the genome of each virus has 5 to 6 components coding for protein other than Rep proteins, which are referred to as non-Rep proteins.

The genus includes species with multiple genomic DNAs that are unidirectionally transcribed, Coconut foliar decay virus (CFDV), a tentative species within the genus, has similar morphology but differs from the assigned members by containing a single circular ssDNA of 1291 nts which is proposed to be transcribed bidirectionally, by having a capsid protein of Mr ca. 24 x $10^3$ and by being transmitted by a plant hopper.

There are a number of species of Nanoviruses including, banana bunchy top virus (BBTV) (S56276, U18077 to U18079, L32166 & L32167, U02312, L32166 & L32167, U02312), Faba bean necrotic yellows virus (FBNY) (X80879, Y11405 to Y11409, (FBNYV) AJ005964 to AJ005968, Milk vetch dwarf virus (MDV) (AB000920 to 000927, 0009046, 0009047), Subterranean clover stunt virus (SCSV) (U16730 to U16736).

Comparison of the 15 potential Rep proteins identified from assigned and tentative members of the genus revealed amino acid identities ranging from about 28% to 90%. In addition, there is variability among Rep proteins, not only from different nanovirus species but also from a given virus. All Rep proteins of the assigned *Nanovirus* species have most of the amino acid domains characteristic of Rep proteins. The nanovirus Rep proteins differ form those of members of the family Geminiviridae in being smaller (about 33 kDa) having a slightly distinct dNTP-binding motif (GP/SXGG/NEGKS/T), lacking the Rb-binding motif, and in sharing amino acid sequence identities of only 17 to 22% with them. Moreover, the assigned species are clearly distinct from the geminiviruses in particle

morphology, genome size, number and size of DNA components, mode of transcription, as well as in vector species. The putative Rep protein of porcine circovirus (family Circoviridae) has only insignificant identity (18 to 23%) with nanovirus Rep proteins, and has a bidirectionally transcribed ssDNA genome resembling that of CFDV. All of these viruses have a conserved nonanucleotide motif at the apex of the stem-loop sequence which is consistent with operation of a rolling circle model for DNA replication. Additional information concerning *Nanoviruses* is found in Boevink, P, *J. Virol 207*:354-361 (1995), Katul, L. *et al.*, *J. Gen. Virol. 79*:3101-3109 (1998), Harding, R. *et al., J. Gen. Virol. 78*:479-486 (1993), Katul *et al., Virology 233*:247-259 (1997).

*Circoviridae*

The *Circovirus* virions exhibit icosahedral symmetry and do not possess an envelope. Ranges of reported virion sizes for *Chicken anemia virus* (CAV), *Porcine circovirus* (PCV) and *Beak and feather disease virus* (BFDV) are 19.1 - 26.5 nm, 17 - 20.7 nm and 12 - 20.7 nm respectively. Diameters of virus particles are approximately 20% greater when negatively stained using uranyl acetate as opposed to the more commonly used phosphotungstate. A capsid structure consisting of 32 hollow morphological subunits arranged in a T = 3 icosahedron has been proposed for CAV. The buoyant density of virions in CsCl is 1.33 Ð 1.37g/cm$^3$. CAV and PCV possess sedimentation coefficients of 91S and 57S respectively. CAV and PCV are resistant to inactivation by treatment at pH 3 and both viruses can withstand incubation at 70¡C for 15 min. CAV is resistant to treatment with chloroform and ether; PCV is resistant to treatment with chloroform. Both CAV and PCV are at least partially resistant to sodium dodecyl sulfate. The effects of pH, solvents and detergents on the infectivity of BFDV are not known due to the lack of an *in vitro* culture system..

The virions contain circular ssDNA. The genomes of CAV and PCV contain 2,298 and 1,759 bases respectively. The BFDV genome is about 2.0 kb in size. The CAV genome is of negative sense. Information concerning the sense of the PCV and BFDV genomes has not been reported. The PCV genome contains a

nonanucleotide sequence motif (TAGTATTAC), which is found at the apex of a potential stem loop and which is identical or highly similar to those found in bacterial and plant viruses with circular, ssDNA genomes (*Microviridae*, *Nanovirus* and *Geminiviridae*).

CAV and PCV virions contain one protein, Mr = 50 and 36 x $10^3$ respectively. BFDV is reported to contain three proteins, Mr = 26.3, 23.7, and 15.9 x $10^3$. CAV has two non-structural proteins, Mr = 24 x $10^3$ (VP2) and 13.6 x $10^3$ (VP3), the smaller of which causes apoptosis *in vitro*. The non-structural proteins of PCV have not been characterized. The N-terminal of the CAV CP shares homologies with histone proteins consistent with it having a DNA-binding role within the virion. The protein (Mr = 35.7 x $10^3$) encoded by the largest PCV ORF has amino acid sequence homology with the replication-associated proteins of plant viruses with circular, ssDNA genomes.

CAV and PCV DNA replicate using circular ds replicative form (RF) DNAs. Nucleic acid and protein homologies shared with plant geminiviruses are consistent with PCV DNA replicating by a rolling circle mechanism. The origin of replication of PCV DNA has been mapped.Only one strand of the CAV RF is transcribed to produce a polycistronic messenger RNA ( ~ 2.1 kb) which contains 3 partially overlapping ORFs encoding proteins of Mr = 52 x $10^3$ (VP1, CP), 24 x $10^3$ (VP2) and 13.6 x $10^3$ (VP3). All three proteins are detected in electron dense bodies within the nuclei of virus-infected cell cultures. CAV particles have not been directly observed within infected cells. The PCV RF sequence has 6 ORFs larger than 200 nts, which occur in both positive and negative sense orientations.

CAV causes transient anemia and immunosuppression in baby chicks. BFDV causes chronic and ultimately fatal disease in large psittacine birds. PCV-like viruses have been associated with a recently described condition of pigs known as post-weaning multisystemic wasting syndrome. Cells of the hematopoietic system are infected by CAV and BFDV At present all assigned members of the family have been classified within a single genus. However differences in virion size,

genome size and genome organization may provide the basis for definition of more than one genus in the future.

*Circoviruses* include *Beak and feather disease virus* (BFDV), Chicken anemia virus (M55918, M81223)(CAV) and *Porcine circovirus* (U49186, Y09921) (PCV). PCV is similar to members of the families, *Geminiviridae* and *Microviridae*, in that it exhibits nucleic acid and protein homologies related to rolling circle DNA replication. The animal circoviruses are similar to plant nanoviruses such as *Banana bunchy top virus*, *Coconut foliar decay virus* and *Subterranean clover stunt virus*, which possess non-enveloped, icosahedral capsids (18 - 20 nm in diameter) and circular, ssDNA genomes (0.85 - 1.3 kb in size). These plant viruses, formerly regarded as Òunassigned viruses in the family *Circoviridae*Ó, are now classified in an unassigned genus *Nanovirus*. Further information cocnerning *circoviruses* can be found in Mankertz, A *et al.*, *J. Virol. 71*:2562-2566 (1997), Tischer, *et al.*, *Nature 295*:64-66 (1982), Todd, D. *et al.*, *Arch. Virol. 117*:129-135 (1991).

There is a need for technology aimed at controlling viruses such as geminiviruses or any virus that can be inhibited by a Rep-iteron antagonist. Basically, there are two approaches. A first approach is altering the movement of the virus in the infected plant or animal and the second approach is shutting down the virus replication. In addition there are added complications concerning such viruses. These are that the viruses, are difficult to distinguish from each other with a complicated taxonomy, often infect plants in mixtures and recombine frequently (Padidam *et al.*, *Virol. 265*:218-225, (1999) and the time frame for these recombinations is unknown. Therefore, there are several issues associated with virus infection: the precise and proper molecular identification of field isolate, the development of one or more methods to control by biotechnology infection on any and all crops by geminiviruses where needed, and the development of a method that will either be general (or non-specific) or alternatively, be very specific but exploitable in such way that it would become non-specific. The invention meets many of these needs.

## *Summary of the Invention*

The invention is directed to a method for producing resistance in a plant to a geminivirus comprising introducing a geminivirus replication associated protein (Rep)-iteron antagonist into a plant, plant cell or propagule, wherein said antagonist is selected from the group consisting of a nucleotide sequence of a geminivirus iteron capable of binding to a Rep protein and a defective Rep, wherein said defective Rep comprises a conserved geminivirus iteron binding site. Preferably, the invention is directed to the use of sequences found in Figs (SEQ ID NOS:1-8, 10-35, 37-41, 43-49, 54-57, 59-62, 64-98). Embodiments of the invention are drawn to use of a Rep protein that may form a dimer with a wild-type geminivirus Rep protein or where the Rep protein comprises from two to thirty different or the same conserved iteron binding sites.

Another embodiment of the invention is directed to use in the above method of a defective replication associated protein (Rep) selected from the group consisting of truncated geminivirus Rep protein, a modified Rep protein capable of binding a geminivirus iteron sequence, or a Rep protein fragment capable of binding a geminivirus iteron sequence.

The invention is further directed to a vector containing a nucleotide sequence that encodes a defective geminivirus replication associated protein, wherein said encoded protein comprises a polypeptide having an amino acid sequence of a conserved geminivirus iteron binding site or a mutant thereof. Preferably the vector is expressed in plants. The vector also preferably encodes a polypeptide comprising a sequence as shown in Figs. 1A-1C (SEQ ID NO:1-8, 10-35, 37-41, 43-49, 54-57, 59-62, 64-98). An embodiment of the invention is directed to a polypeptide that forms a dimer with a wild-type geminivirus Rep protein. or one that comprises from two to thirty different conserved iteron binding sites. Another embodiment of the invention comprises a nucleotide sequence encoding at least two different Rep proteins.

The invention is also directed to a nucleic acid molecule containing a nucleotide sequence comprising an isolated conserved geminivirus iteron. A

preferable embodiment comprises an isolated DNA sequence comprising GGTGTCTGGAGTC (SEQ ID NO:111).

The invention is further directed to compositions for producing resistance to a geminivirus in plants comprising the vectors or nucleic acids of any of the embodiments of the invention.

Another aspect of the invention is directed to transgenic plants, cells, propagules or seeds that comprise any of the vectors or nucleic acids of the invention. A preferred embodiment comprises a nucleic acid molecule having a nucleotide sequence comprising a conserved geminivirus iteron.

The invention is further directed to an isolated nucleic acid molecule comprising a nucleotide sequence for a conserved geminivirus iteron. A preferable embodiment of the invention may be directed to a nucleotide sequence comprising at least two geminivirus iterons. Another embodiment of the invention is directed to a nucleotide sequence that comprises from two to thirty different classes of geminivirus iteron shown in Figs. 1A-1C.

The invention is also directed to a truncated Rep protein. Preferably the truncated Rep protein may be an isolated polypeptide selected from the group consisting of $AC1_{1-21}$, $AC1_{1-60}$, $AC1_{1-52}$, $AC1_{1-111}$, $AC1_{1-114}$, $AC1_{1-160}$ and $AC1_{1-180}$. or a nucleic acid encoding said polypeptides. In another preferable embodiment the truncated protein comprises at least $AC1_{1-160}$.

A further aspect of the invention is directed to a method for inhibiting geminivirus replication in a plant comprising introducing a geminivirus replication associated protein (Rep)-iteron antagonist into said plant, said antagonist selected from the group consisting of a nucleotide sequence defining a geminivirus iteron capable of binding to a Rep protein and a defective Rep protein, wherein said defective Rep comprises a conserved geminivirus iteron binding site.

The invention is further drawn to a method for providing resistance to infection by geminiviruses in a susceptible plant comprising: a) transforming susceptible plant cells with a DNA molecule that comprises operatively linked in sequence in the 5' to 3' direction i) a promoter region that functions in plant cells to cause the production of an RNA sequence; and ii) a gene encoding a defective

Rep protein, wherein said defective Rep protein comprises a conserved geminivirus iteron binding site; said method further comprising b) selecting said plant cells that have been transformed; c) regenerating said plant cells to provide a differentiated plant; and d) selecting a transformed plant that expresses said defective Rep gene at a level sufficient to render the plant at least partially resistant to infection by the geminivirus.

The invention is further directed to an at least partially virus-resistant transformed plant normally susceptible to infection by a geminivirus having inserted into its genome a DNA molecule that comprises operatively linked in sequence in the 5' to 3' direction; i) a promoter region that functions in plant cells to cause the production of an RNA sequence; and ii) a gene encoding a defective Rep protein, wherein said defective Rep comprises a conserved geminivirus iteron binding site.

The invention is directed to a method for producing at least partial resistance to a virus or a method for reducing replication of a virus in a plant, plant cell, propagule, animal or animal cell comprising introducing a replication associated protein (Rep)-iteron antagonist into a plant, plant cell, propagule, animal or animal cell, wherein said antagonist is selected from the group consisting of a nucleotide sequence of an iteron capable of binding to a Rep protein and a defective Rep, wherein said defective Rep comprises a conserved iteron binding site, and wherein said Rep-iteron antagonist renders the infected plant, plant cell, propagule, animal or animal cell at least partially resistant to the infection. Preferably, the invention is directed to the use of sequences at least 50% identical to those found in Figs 1A-1C (SEQ ID NOS:1-8, 10-35, 37-41, 43-49, 54-57, 59-62, 64-107). More preferably the sequences are at least 60%, 70%, 80%, 90% 95% or 99% identical. Another preferable embodiment is directed to producing resistance to infection from a *Nanovirus or Circoviridae*. Additional embodiments of the invention are drawn to use of a Rep protein that may form a dimer with a wild-type geminivirus Rep protein or where the Rep protein comprises from two to thirty different or the same conserved iteron binding sites. Another embodiment of the invention is drawn to reducing infection or reducing

DNA replication of any virus that replicates in a manner similar to the geminivirus, i.e. dependent on the binding of a Rep protein to an iteron. Embodiments of the invention are also drawn to use of a Rep protein that may form a dimer with a wild-type geminivirus Rep protein or where the Rep protein comprises from two to thirty different or the same conserved iteron binding sites.

The invention is further directed to a composition for producing at least partial resistance to a virus or for reducing replication of a virus in a plant, plant cell, propagule, animal or animal cell wherein said composition is used in any of the above methods. The compositions may also include solutions that are physiologically compatible with the organism of interest.

The invention is also directed to a Rep-iteron antagonist comprising a nucleic acid sequence encoding a Rep protein or fragment thereof that binds to an iteron wherein viral infection or DNA replication of the virus causing the infection is reduced following said antagonist binding to an iteron..

The invention is further directed to a Rep-iteron antagonist comprising a nucleic acid sequence that competes for binding of a Rep protein with the iteron of the virus causing the infection, wherein viral infection or DNA replication of the virus causing the infection is reduced following said antagonist binding to the Rep protein. Another embodiment of the invention is directed to a Rep-iteron antagonist comprising a polypeptide or the nucleic acid sequence encoding a polypeptide comprising the sequence FLTY or KAYIDK.

The invention is further directed to a Rep-iteron antagonist comprising a polypeptide or the nucleic acid sequence encoding a polypeptide selected from the group consisting of: FLTYPqC wherein q is a basic or a polar amino acid, HIHxUUQ wherein U is a bulky hyrophobic residue and xxYxxK wherein x may be any amino acid.

The invention is also directed to a vector comprising a nucleic acid sequence encoding any of the Rep-iteron antagonists of the invention.

## *Brief Description of the Figures*

***Figs. 1A-1C.*** **Sequences of Iterons and Rep N-terminal Sequences.** Figure 1A - 1B. Begomoviruses (SEQ ID NOs:1-8 10-41, 76-98, 99-100, 108-110) Figure 1C - Mastreviruses (SEQ ID NO:43-49, 54-69, 101-105), Curtoviruses (SEQ ID NO:70-74,106-107) and Topcuvirus (SEQ ID NOS: 75, 107).

***Fig. 2A-2C.*** **Immuno-precipitation of Rep protein from crude lysates of Sf9 cells using anti AC1 antibody.**

***Fig. 2A.***. The Rep proteins from the severe strain (A1, lane 1) and the mild strain (A2, lane 2) of ToLC-NdeV were expressed from the polyhedrin promoter of AcNPV in Sf9 cells and detected using anti AC1 polyclonal antiserum.

***Fig. 2B.*** Coomassie blue stained gel showing the purified Rep proteins from the severe (lane 2) and mild (lane 5) strains of ToLC-NdeV. Lane 1 represents the marker and lane 4 shows the crude lysate from the pellet fraction.

***Fig. 2C.*** Western blot using a polyclonal anti-AC1 antiserum. Stepwise eluates of the purified protein were collected from the $Ni^{2}$ affinity column in 20mM Tris, 500mM NaCl a 500mM imidazole (pH 7.9) and detected using the anti AC1 antiserum. Lanes 1-4 represent stepwise aliquots of the purified protein of the severe strain and the lanes 5-7 show similar fractions of the protein purified from the mild strain of ToLC-NdeV.

***Fig. 3. Electrophoretic mobility shift assays showing the interaction of the Rep protein of severe (A1) and the mild (A2) strains of ToLC-NdeV with different common region fragments.*** CR-s and CR-m refer to the 52 bp common region fragment derived from the intergenic region of the viral DNA. bs-s and bs-m denote the 13-bp repeat motifs in the common region of the severe and the mild strain respectively. $^{32}P$ labeled DNA fragments (CR or bs)

were incubated in the presence (+) or absence (-) of competitor DNA to test the specificity of binding. All reactions contained 200 ng of poly dI.dC and were analyzed on 4% polyacrylamide gels. The reactions in lanes 3 and 8 contained 50x molar excess of appropriate, unlabelled 13-bp DNA as the specific competitor and lanes 4 and 9 show the complex formation in the presence of 1000X molar excess of non-specific competitor (pUC 18) DNA. The Rep proteins of the two strains did not bind to heterologous binding site sequences as seen in lanes 5 and 10.

***Figs.4A-4B. DNA sequence requirements for binding by the Rep protein.***

***Fig. 4A.*** Labeled synthetic oligonucleotides with variations in the sequence and arrangement of iterons were used as probes to analyze their effect on binding by the Rep protein of severe (lanes 1-6) and mild (lanes 7 to 10) strains of ToLC-NdeV. The key to the sequence of iterons used as probes is as follows: IT 1/2 (5′ severe, 3′ mild, lanes 1 and 7); I T3/4 (5′unrelated, 3′ severe, lanes 2 and 8); IT 5/6 (5′ unrelated, 3′mild, lanes 5 and 9); IT 7/8 (5′mild, 3′unrelated, lanes 6 and 10); IT 9/10 (5′ severe repeated, lane 3); IT 11/12 (3′ severe repeated, lane 4).

***Fig. 4B.*** Labeled synthetic oligonucleotides with variations in the spacing and number of iterons were used as probes to analyze their effect on binding by the Rep protein. The key to the sequence of iterons used as probes is as follows: IT 13/14 (spacing within the iterons is increased to 6 nucleotides, lanes 1 and 3); IT 15/16 (no spacing between the iterons, lanes 2 and 4); IT 17/18 (5′monomer iteron, lanes 5 and 7); IT 19/20 (5′monomer repeated 4 times, lanes 6 and 8). Lane 9 shows the free probe.

***Fig. 5. Replication of Rep protein binding site mutants.*** Plasmids (2μg) containing the viral replicons mutated at their binding site sequence in the origin were electroporated into tobacco protoplasts. Total DNA was isolated 48h after transfection, resolved on agarose gels and analyzed by

Southern hybridization using $^{32}$P-labelled AC1 DNA fragment (nts 2113 to 2695) as a probe. The single stranded (ss) and the supercoiled (sc) forms of the viral DNA are indicated. The virus mutants were given identical names as the oligonucleotides used to alter their iteron sequence for the sake of convenience. The key to the mutants is as follows: Lanes 2 and 13 (IT 1/2); lanes 3 and 14 (IT 3/4); lanes 4 and 15 (IT 5/6); lanes 5 and 16 (IT 7/8); lane 6 (IT 9/10); lane 7 (IT 11/12); lanes 8 and 17 (IT 13/14); lanes 9 and 18 (IT 15/16); lanes 10 and 19 (IT 17/18); lanes 11 and 20 (IT 19/20). Lanes 1 and 12 represent the wild type controls for the severe and the mild strain of ToLC-NdeV respectively.

***Fig. 6A-6B.*** **Accumulation of Viral DNA in BY-2 Protoplasts.** The protoplasts were transfected with truncated and full length Rep proteins.

***Fig. 7.*** **Replication of Truncated Rep protein in N. *benthamiana* Plants.**

***Figs. 8A-8C.*** **Genome organization of tomato leaf curl virus from New Delhi, (ToLCV-Nde).** This figure shows mutations made in Rep gene and in the common region. **(Fig. 8A)** Genome maps of DNA-A and DNA-B of severe strain. The genes encoding conserved proteins in geminiviruses are shown as solid arrows. Rep, TrAP, REn and CP on DNA-A represent the replicase associated protein (AC1), the transcriptional activator protein (AC2), the replication enhancer (AC3) and the coat protein (AV1) respectively. The MP and NSP on DNA-B are the movement protein (BC1) and the nuclear shuttle protein (BV1) respectively. The genome organization of DNA-A of both severe and the mild strain are identical. Relevant restriction sites used for mutagenesis are indicated. **(Fig. 8B)** Schematic representation of mutants made in Rep gene of mild and severe strain DNA-A. Fragments were exchanged at the N-(*Nco* I to *Xba* I) or C-(*Cla* I to *Cla* I) terminal of Rep gene between the strains. The ToLCV severe strain is indicated in white hatched lines whereas the mild strain is shown by black

lines. **(Fig. 8C)** A schematic showing organization of origin of replication in geminiviruses (not to scale). The mutations made in the putative binding site of Rep protein and the N-terminal region of Rep gene are shown. The hairpin, TATA box and the major ORFs in virus sense and complementary sense are indicated. The repeat sequence forming the binding site is shown as two solid arrows near the TATA box. The putative binding sites identified for the severe strain DNA-A and DNA-B and the mild strain DNA-A are indicated. Substitution mutations made in the N-terminal of Rep gene of mild and the severe strain together with point mutations made in the Rep protein binding sites are presented. (SEQ ID NO:112-120) The panel on the left shows the sequence of first ten amino acids on the Rep protein of Al and A2 starting with the initiation codon, methionine (M), while the middle panel indicates the putative binding site sequence on the corresponding mutants (indicated on the right). (SEQ ID NOS: 121-129)

***Fig. 9.*** **Southern blot analysis of viral DNA in *N. tabacum* protoplasts inoculated with different mutants of ToLCV**. Total DNA was extracted from protoplasts 48h after transfection and electrophoresed through 1% agarose gel without ethidium bromide, transferred to nylon membrane and hybridized with $^{32}p$ labelled DNA-A and DNA-B specific probes. Panel A shows replication ability of mutants made in severe strain DNA-A and probed with A-component (lanes 1-9) and B-component (lanes 10-18) specific probes. Panel B shows replication efficiency of mutants made in the mild strain DNA-A probed with A-component (lanes 1-9) and B-component (lanes 10-18) specific probes. The positions of single stranded (ss) and supercoiled (sc) viral DNA are indicated. Each lane contains 4 μg of DNA obtained from protoplasts in a single transfection.

***Fig. 10.*** **Southern blot analysis of viral DNA in *N. benthamiana* plants inoculated with ToLCV mutants**. Total DNA was extracted from newly emerging leaves three weeks after bombardment and electrophoresed in 1% agarose gels without ethidium bromide, transferred to nylon membrane and

hybridised with $^{32}$p labeled DNA-A and DNA-B specific probes. Panel A shows replication competence of severe strain mutants probed with A-component (lanes 1-7) and B-component (lanes 8-14) specific probes. Panel B shows replication efficiency of mutants made in the mild strain DNA-A and probed with A-component (lanes 1-9) and B-component (lanes 10- 18) specific probes. The position of single stranded (ss) and supercoiled (sc) viral DNA are indicated.

## *Detailed Description of the Preferred Embodiments*

It has been discovered that the geminivirus replication associated protein (hereafter called Rep protein or "Rep") specifically recognizes and binds to short stretches of geminivirus DNA sequences called binding sites or iterons, and this binding marks the first step in the replication of the virus in plants. These results are elaborated in the Examples of the specification that provide detailed descriptions of iterons and interactions with Rep proteins.

*Definitions*

In order to provide a clearer understanding of the specification and claims, the following definitions are provided.

***Amino Acid Sequences*** - The amino acid sequences herein use either the single letter or three letter designations for the amino acids. These designations are well known to one of skill in the art and can be found in numerous readily available references, such as for example in *Cooper, G.M., The Cell* 1997, ASM Press, Washington, D.C. or Ausubel et al., *Current Protocols in Molecular Biology,* 1994 (Also *see* 37 C.F.R. § 1.821).

***Cloning vector:*** A plasmid or phage DNA or other DNA sequence which is able to replicate autonomously in a host cell, and which is characterized by one or a small number of restriction endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion, and into which a DNA fragment may be spliced in order to bring about its replication and cloning. The cloning vector may further contain a marker suitable for use in the identification of cells

transformed with the cloning vector. Markers, for example, provide tetracycline resistance or ampicillin resistance.

***DNA construct***. As used herein, "DNA construct" should be understood to refer to a recombinant, man-made DNA, either linear or circular.

***Derivative or Functional Derivative:*** The term "derivative" or "functional derivative" is intended to include "variants," the "derivatives," or "chemical derivatives" of the Rep molecule. A "variant" of a molecule or derivative thereof is meant to refer to a molecule substantially similar to either the entire molecule, or a fragment thereof. An "analog" of a molecule or derivative thereof is meant to refer to a non-natural molecule substantially similar to either the Rep molecules or fragments thereof. Chemical and functional derivatives of the Rep protein are considered embodiments of the application.

Rep derivatives contain changes in the polypeptide relative to the native Rep polypeptide of the same size. A molecule is said to be "substantially similar" to another molecule if the sequence of amino acids in both molecules is substantially the same, and if both molecules possess a similar biological activity. Thus, two molecules that possess a similar activity, may be considered variants, derivatives, or analogs as that term is used herein even if one of the molecules contains additional amino acid residues not found in the other, or if the sequence of amino acid residues is not identical. Rep derivatives, however, need not have substantially similar biological activity to the native molecule.

As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, strength, specificity, affinity, biological half-life, etc. The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, etc. and will be apparent to those of ordinary skill in the art. "Functional derivatives" include those polypeptides that bind iteron sequences and those nucleic acid sequences that bind Rep proteins.

***Expression vector***. As used herein, an "expression vector" is a DNA construct that contains a structural gene operably linked to an expression control

sequence so that the structural gene can be expressed when the expression vector is transformed into an appropriate host cell. Two DNA sequences (such as a promoter region sequence and a sequence encoding a Rep derivative) are said to be "operably linked" if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region sequence to direct the transcription of the desired sequence, or (3) interfere with the ability of the desired sequence to be transcribed by the promoter region sequence. Thus, a promoter region would be operably linked to a desired DNA sequence if the promoter were capable of effecting transcription of that DNA sequence.

***Fragment:*** A "fragment" of a molecule is meant to refer to any polypeptide subset of these molecules. A truncated Rep may considered to be a fragment of the whole molecule

***Fusion protein:*** By the term "fusion protein" is intended a fused protein comprising a protein or polypeptide either with or without a "selective cleavage site" linked at its N-terminus, which is in turn linked to an additional amino acid leader polypeptide sequence.

***% Identity:*** Whether any two polypeptides or polynucleotides are for example, at least 90% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA 85*:2444 (1988). Alternatively the BLAST function of the National Center for Biotechnology Information database may be used to determine identity

The terms homology and identity are often used interchangeably. In this regard, percent homology or identity may be determined by methods known to those of skill in the art. For example, by comparing sequence information using a GAP computer program, version 6.0, available from the University of Wisconsin Genetics Computer Group (UWGCG). The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443 (1970), as revised by Smith and Waterman (*Adv. Appl. Math.* 2:482 (1981). Briefly, the GAP program defines similarity as the number of aligned symbols (i.e.,

nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences.

In general, sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be calculated using published techniques. (See, *e.g.*: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math 48*:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math 48*:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., *et al.*, *Nucleic Acids Research 12(I)*:387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., *et al.*, *J Molec Biol 215*:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. More specifically, a test polypeptide may be defined as any polypeptide that is 90% or more identical to a reference polypeptide. As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of 100 amino acids, that no more than 10% (i.e., 10 out of 100) amino acids in the

test polypeptides differ from that of the reference polypeptides. Such differences may be represented as point mutations randomly distributed over the entire length of the amino acid sequence of the invention or they may be clustered in one or more locations of varying length up to the maximum allowable, e.g. 1/14 amino acid difference (approximately 90% identity). Differences are defined as amino acid substitutions, or deletions. Embodiments of the claimed invention may include those polypeptides or nucleic acid sequences that are at least 90% identical to specifically claimed sequences.

***Isolated:*** A term meaning altered from the natural state. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated," as the term is employed herein. Thus, a polypeptide or polynucleotide produced and/or contained within a recombinant host cell is considered isolated for purposes of the present invention. Also intended as an "isolated polypeptide" or an "isolated polynucleotide" are polypeptides or polynucleotides that have been purified, partially or substantially, from a recombinant host cell or from a native source. For example, a recombinantly produced version of a compound or derivatives thereof can be substantially purified by the one-step method described in Smith and Johnson, *Gene 67:*31-40 (1988). The terms isolated and purified are sometimes used interchangeably.

By "isolated DNA" is included DNA free of the coding sequences of those genes that, in the naturally-occurring genome of the organism (if any) from which the DNA of the invention is derived, immediately flank the gene encoding the DNA of the invention. The isolated DNA may be single-stranded or double-stranded, and may be genomic DNA, cDNA, recombinant hybrid DNA, or synthetic DNA. It may be identical to a native DNA sequence, or may differ from such sequence by the deletion, addition, or substitution of one or more nucleotides. Isolated single-stranded DNAs of the invention may be detectably labeled for use as hybridization probes, and may be antisense.

Isolated or purified as it refers to preparations made from biological cells or hosts should be understood to mean any cell extract containing the indicated DNA or protein including a crude extract of the DNA or protein of interest. For example, in the case of a protein, a purified preparation can be obtained following an individual technique or a series of preparative or biochemical techniques and the DNA or protein of interest can be present at various degrees of purity in these preparations. The procedures may include for example, but are not limited to, ammonium sulfate fractionation, gel filtration, ion exchange change chromatography, affinity chromatography, density gradient centrifugation and electrophoresis.

A preparation of DNA or protein that is "pure" or "isolated" should be understood to mean a preparation free from naturally occurring materials with which such DNA or protein is normally associated in nature. "Essentially pure" should be understood to mean a "highly" purified preparation that contains at least 95% of the DNA or protein of interest.

A cell extract that contains the DNA or protein of interest should be understood to mean a homogenate preparation or cell-free preparation obtained from cells that express the protein or contain the DNA of interest. The term "cell extract" is intended to include culture media, especially spent culture media from which the cells have been removed.

***Iteron***: The term "iteron" refers to a direct repeat motif of 6-12 bp within the common region of the viral genome between the TATA box and the start site for the transcription of for example, the AC1 gene. Iterons have been proposed to serve as the high affinity binding sites of the Rep protein and therefore function as origin recognition sequences.

***Plant:*** The term "plant" should be understood as referring to a multicellular differentiated organism capable of photosynthesis including angiosperms (monocots and dicots) and gymnosperms.

***Plant cell:*** The term "plant cell" should be understood as referring to the structural and physiological unit of plants. The term "plant cell" refers to any cell which is either part of or derived from a plant. Some examples of cells

encompassed by the present invention include differentiated cells that are part of a living plant; differentiated cells in culture; undifferentiated cells in culture; and the cells of undifferentiated tissue such as callus or tumors.

***Plant cell progeny:*** The term "plant cell progeny" should be understood as referring to any cell or tissue derived from plant cells including callus; plant parts such as stems, roots, fruits, leaves or flowers; plants; plant seed; pollen; and plant embryos.

***Propagules*:** The term "propagules" should be understood as referring to any plant material capable of being sexually or asexually propagated, or being propagated *in vivo* or *in vitro*. Such propagules preferably consist of the protoplasts, cells, calli, tissues, embryos or seeds of the regenerated plants.

***Polynucleotide:*** This term generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications have been made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

***Polypeptide:*** Polypeptide, protein and peptide are used interchangeably. The term polypeptide refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, *i.e.*,

peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids and include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in the research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications.

Polypeptides may be branched and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translational modifications or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. *See*, for instance, *Proteins-Structure and Molecular Properties*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993 and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in *Posttranslational Covalent Modification of Proteins*, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Methods in Enzymol. 182*:626-646

(1990) and Rattan *et al.*, "Protein Synthesis: Posttranslational Modifications and Aging", *Ann NY Acad Sci 663*:48-62 (1992).

***Promoter:*** A DNA sequence generally described as the 5′ region of a gene, located proximal to the start codon. The transcription of an adjacent gene(s) is initiated at the promoter region. If a promoter is an inducible promoter, then the rate of transcription increases in response to an inducing agent. In contrast, the rate of transcription is not regulated by an inducing agent if the promoter is a constitutive promoter. Examples of promoters include, but are not limited to the CMV promoter (InVitrogen, San Diego, CA), the SV40, MMTV, and hMTIIa promoters (U.S. Pat. No. 5,457,034), the HSV-1 4/5 promoter (U.S. Pat. No. 5,501,979), the early intermediate HCMV promoter (WO92/17581), ubiquitin, actin, phenylammonia lyase (PAL), CaMV 35S, CsVMW, and RTBV promoters. Also, tissue-specific enhancer elements may be employed. Additionally, such promoters may include tissue and cell-specific promoters of an organism.

***Recombinant Host:*** According to the invention, a recombinant host may be any prokaryotic or eukaryotic host cell which contains the desired cloned genes on an expression vector or cloning vector. This term is also meant to include those prokaryotic or eukaryotic cells that have been genetically engineered to contain the desired gene(s) in the chromosome or genome of that organism. For examples of such hosts, *see* Sambrook *et al., Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989). Preferred recombinant hosts are eukaryotic cells transformed with the DNA construct of the invention. More specifically, mammalian cells are preferred.

***Rep-iteron antagonist***: The term "Rep-iteron antagonist" refers to a defective Rep protein or single or multiple iteron nucleotide sequences that are effective at inhibiting replication of any geminivirus that infects plants.

***Selective cleavage site:*** The term "selective cleavage site" refers to an amino acid residue or residues which can be selectively cleaved with either chemicals or enzymes in a predictable manner. A selective enzyme cleavage site

is an amino acid or a peptide sequence which is recognized and hydrolyzed by a proteolytic enzyme. Examples of such sites include, without limitation, trypsin or chymotrypsin cleavage sites.

***Stringent Hybridization.*** As used herein "stringent hybridization" conditions should be understood to be those conditions normally used by one of skill in the art to establish at least a 95% homology between complementary pieces of DNA or DNA and RNA.

There are only three requirements for hybridization to a denatured strand of DNA to occur. (1) There must be complementary single strands in the sample. (2) The ionic strength of the solution of single-stranded DNA must be fairly high so that the bases can approach one another; operationally, this means greater than 0.2M. (3) The DNA concentration must be high enough for intermolecular collisions to occur at a reasonable frequency. The third condition only affects the rate, not whether renaturation/hybridization will occur.

Conditions routinely used by those of skill in the art are set out in readily available procedure texts, *e.g.*, Ausubel. F. *et al.*, *Current Protocols in Molecular Biology*, Vol. I, Chap. 2.10, John Wiley & Sons, Publishers (1994) or Sambrook *et al.*, *Molecular Cloning*, Cold Spring Harbor (1989), the entire contents of which are incorporated herein by reference. As would be known by one of skill in the art, the ultimate hybridization stringency reflects both the actual hybridization conditions as well as the washing conditions following the hybridization, and one of skill in the art would know the appropriate manner in which to change these conditions to obtain a desired result.

For example, a prehybridization solution should contain sufficient salt and nonspecific DNA to allow for hybridization to non-specific sites on the solid matrix, at the desired temperature and in the desired prehybridization time. For example, for stringent hybridization, such prehybridization solution could contain 6x sodium chloride/sodium citrate (1xSSC is 0.15 M NaCl, 0.015 M Na citrate; pH 7.0), 5x Denhardt's solution, 0.05% sodium pyrophosphate and 100 µg per ml of herring sperm DNA. An appropriate stringent hybridization mixture might then

contain 6x SSC, 1x Denhardt's solution, 100 μg per ml of yeast tRNA and 0.05% sodium pyrophosphate.

Alternative conditions for DNA-DNA analysis could entail the following:

1) prehybridization at room temperature and hybridization at 68°C;
2) washing with 0.2x SSC/0.1% SDS at room temperature;
3) as desired, additional washes at 0.2x SSC/0.1% SDS at 42°C (moderate-stringency wash); or
4) as desired, additional washes at 0.1x SSC/0.1% SDS at 68°C (high stringency).

Known hybridization mixtures, *e.g.*, that of Church and Gilbert, *Proc. Natl. Acad.. Sci. USA 81*:1991-1995 (1984), comprising the following composition may also be used: 1% crystalline grade bovine serum albumin/1mM EDTA/0.5M $NaHPO_4$, pH 7.2/7% SDS. Additionally, alternative but similar reaction conditions can also be found in Sambrook *et al.*, *Molecular Cloning*, Cold Spring Harbor (1989). Formamide may also be included in prehybridization/ hybridization solutions as desired. The invention may include DNA sequences that stringently hybridize to specifically disclosed sequences of the invention.

***Transgenic plant***: The term "transgenic plant" should be understood as referring to a plant having stably incorporated exogenous DNA in its genetic material. The term also includes exogenous DNA which may be introduced into a cell or protoplast in various forms, including, for example, naked DNA in circular, linear or supercoiled form, DNA contained in nucleosomes or chromosomes or nuclei or parts thereof, DNA complexed or associated with other molecules, DNA enclosed in liposomes, spheroplasts, cells or protoplasts. Transgenic plants are also considered to include at least the progeny of such created plants that express the transgene originally inserted into the first generation plant.

It should be understood that these descriptions are not meant to be definitive or limiting and may be adjusted as required by those of ordinary skill in the art to accomplish the desired objective.

It has also been discovered that the iteron binding domain is located in the amino-terminus portion of the Rep protein, and in particular, conserved and variable domains in the amino terminus of the Rep protein have been discovered which allow binding to the iteron, and the variable domain maps to residues 1-10 of the Rep protein which defines the sequence specificity of the iteron binding site. More importantly, it has been discovered that iteron sequences are conserved around residues 11-15 of the Rep protein among large numbers of isolates of geminivirus (see, e.g. Figs. 1A-1C) such that the Rep protein from one virus can bind the iteron of another virus. A small number of different iteron sequences have been identified which encompass all known geminiviral iteron sequences.

The binding of Rep to its cognate site is therefore now known to be sequence-specific and the efficiency of the binding is related to the sequence of the first 10-12 aa of the Rep protein and to the sequence of the iterons (i.e., a pair of typically 5 nucleotides separated by two to twenty spacer nucleotides). See Figure 1 for the structure and sequence of an iteron and Rep N-terminal sequences.

In vivo experiments are described in the examples in which point mutations in both the Rep and the iterons are made to demonstrate a correlation between these two entities. Furthermore, in vitro DNA binding experiments described in the examples demonstrates the physical interaction between the two entities. In addition, in vivo data described in the examples demonstrates that different wild type and truncated Reps compete in binding to viral iteron sequences and compete in their influence upon virus replication.

The data presented in the Examples (1) explains the principle of replication specificity of geminiviruses, (2) defines the minimum sequences of Rep that are essential for specific recognition between related strains in geminiviruses, (3) establishes the existence of iterons and their role in binding, and (4) provides the molecular basis whereby the two elements, Rep and iterons, can be manipulated to control geminivirus replication. Such explanations, however, are not meant to limit the scope of the invention.

Examples of molecular control include expressing truncated Reps in single or multiple combinations, each of which bind to one or more geminivirus iterons,

and thereby inhibit viral replication, or expressing nucleotide sequences which contain one or more geminivirus iteron sites that would bind and trap the wild type Rep protein, and thereby interfere with geminivirus replication.

Figs. 1A-1C are a table which shows the amino acid residue sequence of the iteron binding domain of a large number of Rep proteins from different geminivirus isolates, shows the conserved nature of the binding site domain by arranging the geminvirus Rep protein in different conserved groups, and illustrates the conserved relationship between the different geminivirus isolates and the known iterons. Figs. 1A-1C also define iteron "classes" which distinguish the Rep protein binding specificity such that a Rep protein from one iteron class will not bind the iteron from another iteron class. Because one or more iteron sequences may be targeted by the present invention, the invention describes in one embodiment the use of a single iteron; in another embodiment, the use of two or more iterons, preferably about two to about thirty iterons, although the number can vary widely in view of the fact that Figs. 1A-1C illustrate the iteron sequence for about 28 isolates and because it is likely that additional field isolates will be cloned and sequenced.

The invention therefore describes a method for producing in a plant resistance to infection by a geminivirus. The method can be practiced using a variety of approaches based on the basic scientific finding that a conserved iteron binding site is located in the amino terminus of a Rep protein.

The invention describes in one embodiment a method for producing in a plant resistance to a geminivirus comprising introducing a geminivirus replication associated protein (Rep)-iteron antagonist into said plant, where the antagonist is selected from the group consisting of (1) a nucleotide sequence defining a geminivirus iteron capable of binding to a Rep protein, and (2) a defective Rep protein, wherein the defective Rep comprises a conserved geminivirus iteron binding site. As an antagonist, the functional viral Rep protein competes with the viral iteron sequence for binding and therby inhibits viral replication. A Rep-iteron antagonist can be used in a variety of methods and compositions according to the invention.

In one embodiment for practicing the method, one can introduce into a plant a defective Rep protein which binds viral iterons and, upon binding to the iteron, blocked by competition the binding of wild-type Rep proteins, thereby inhibiting viral replication. The defective Rep protein can be any of a variety of polypeptides which possess the ability to bind a geminivirus iteron sequence, but lack any of a variety of other functions required for geminivirus replication, such as the nicking site activity, the NTP binding site, the AC3 protein interaction site, and the like which render the defective Rep protein incapable of supporting replication. The Rep protein functions to be deleted or mutated are located in the wild-type protein's carboxy-terminal region, whereas the iteron-binding domain has been discovered to map to the amino-terminal portion of the wild-type Rep protein.

Thus, a preferred defective Rep protein of this invention is a truncated Rep protein which contains the amino terminus at least amino acid residues 1-52, preferably amino acid residues 1-52. A typical protein contains a binding site shown in Figs. 1A-1C, comprising a Rep amino acid sequence and has at least 25 to 30 amino acid residues. In another embodiment, a preferred Rep protein has amino acid residues 1-52.

More preferably, a defective Rep protein of this invention further has the ability to interact with (i.e., bind to) another Rep protein and form a multimer. The ability of Rep proteins to interact is shown in the Examples, and can be measured by any of a variety of methods, including the interactions as measured herein. In particular, a Rep protein corresponding to amino acid residues 1-56 has the ability to bind an iteron and to bind with defective or wild-type Rep protein.

Thus, the term "defective replication associated protein" or "defective Rep protein" means any of a variety of peptides and proteins including active fragments, fusion proteins containing an active iteron binding site fragment, and derivatives thereof which possess the geminivirus iteron DNA binding activity. Exemplary variations include the deletion mutants and fragmented Rep proteins described in the Exhibits.

A preferred defective Rep protein includes residues 1-56 of the geminivirus Rep protein. In another embodiment, a preferred Rep protein is based on the natural sequence of a tomato leaf curl virus from New Dehli (ToLCV-Nde), which has been described in several strains, particularly the mild and severe isolates.

The complete nucleotide sequence of the DNA-A, including the gene which encodes replication associated protein (Rep), of both mild and severe strains of tomato leaf curl virus from New Dehli (ToLCV-Nde) has been determined. The nucleotide sequence for DNA-A of both strains of ToLCV-Nde is deposited with Genebank having accession numbers of U15015 and U15016 for the severe and mild strains, respectively.

In another embodiment, the invention describes a method for simultaneously inhibiting the infection of plants by a large number of different isolates of geminivirus using a broad spectrum defective Rep protein which binds to a conserved iteron present in different geminivirus isolates. An exemplary "broad spectrum defective Rep protein" comprises an amino-terminal Rep protein amino acid residue sequence shown in Figs. 1A-1C. In a related embodiment, one can use a combination of several different classes of broad spectrum defective Rep proteins, such that one member of each "iteron class" is included in the combination of Rep proteins. An exemplary combination includes a small number of different polypeptides, each of which comprises a different amino acid residue sequence for the iteron binding site selected from the corresponding different iteron binding site classes shown in Figs. 1A-1C. Alternatively, one can use any combination of defective Rep proteins to encompass two or more different iteron classes as defined in Figure 1A-1C.

The recitation of many different iteron sequences and corresponding N-terminal sequences of the Rep protein from different viruses demonstrates that variability at the iteron level in geminiviruses is limited. Even though the viruses may originate in different geographical regions, or have diverse host ranges, some isolates share the same iteron sequence and have certain amino acid sequences in the N-terminal region of the Rep protein which give them a common iteron target sequence. For instance, all geminiviruses with the iteron sequence GGTAC (SEQ.

ID NO: ) have the amino acids FQIN (SEQ. ID NO: ) common between them, other than the motif 1, FLTY (SEQ. ID NO: ) , which is conserved throughout the geminiviridae. For other isolates, the homology is not so remarkable. However, the fourth residue downstream of the conserved Phe (F) residue is almost always identical between different isolates having the same iteron sequence.

Based on the sequences shown in Figs. 1A-1C, it is seen that a variety of iteron binding sites are contemplated by the present invention. Thus, it is contemplated that a Rep protein can comprise any one of the known iteron binding site sequences shown in Figs. 1A-1C, such as an amino acid residue sequence comprising a formula selected from the group consisting of -FRVQ- (SEQ. ID NO:126), -FRVN- (SEQ. ID NO:127), -FRIN-(SEQ. ID NO:128), -FRIQ- (SEQ. ID NO:129), -FRLQ- (SEQ. ID NO:130), -FKVQ- (SEQ. ID NO:131), -FKIY- (SEQ. ID NO:132), -FKIN- (SEQ. ID NO:133), -FRLA- (SEQ. ID NO:134), FRLN- (SEQ. ID NO:135), -LKTN- (SEQ. ID NO:136), -FAIN-(SEQ. ID NO:137), -FRLT-(SEQ. ID NO:138), -FNIN- (SEQ. ID NO:139), -FRVN- (SEQ. ID NO:140), -FSIN- (SEQ. ID NO:141), -FKIY- (SEQ. ID NO:142), -FLIN- (SEQ. ID NO:143), -FQIN-(SEQ. ID NO:144), -FQIY-(SEQ. ID NO:145), -FCIN-(SEQ. ID NO:146) , -FCVN- (SEQ. ID NO:147), -FKLN- (SEQ. ID NO:148), -FSVK- (SEQ. ID NO:149), -FSVN- (SEQ. ID NO:150), -FSIK- (SEQ. ID NO:151) , -FYKK- (SEQ. ID NO:152), -FQIK-(SEQ. ID NO:153), -FQIA-(SEQ. ID NO:154) , -FRLQTKY (SEQ. ID NO:154)- , -FRVYSKY- (SEQ. ID NO:156) and -HRNANT- (SEQ. ID NO:157).

It is also contemplated that the replication associated protein (Rep) may bind to an iteron sequence selected from the group of sequences selected from GGAGAXGGAGA (SEQ ID NO:99), GGTGTXGGTGT (SEQ ID NO:100) , GGTACXGGTAC (SEQ ID NO:107 ), GGGGAXGGGGA (SEQ ID NO:109 ), GGGGGXGGGGG (SEQ ID NO: 110), GGTGCGCCCXGGGCGCACC (SEQ ID NO:101) GCGCCTTCXGAAGGCGCG (SEQ ID NO:102) GGTTTGCGXCGCAAACC (SEQ ID NO:103) GGAGGTGCGTCCX-CCTCCACGGG (SEQ ID NO:105), GGAGTXGGAGT (SEQ ID NO:106 ) and GGTACXGGTAC. (SEQ ID NO:107), GTGAGTGXCACTCAC (SEQ. ID NO:

9), GGTACXGGTAC (SEQ. ID NO:36), GGGGAXGGGGA (SEQ. ID NO:42), GGGGGXGGGGG (SEQ. ID NO:50) wherein "X" is 3-30 nucleotides. Alternatively, the replication associated protein (Rep) may bind to a DNA sequence comprising GGTGTCTGGAGTC (SEQ ID NO: 111).

It is also contemplated that the Rep protein can have a modified amino acid residue sequence which comprises a modified (mutated or altered) iteron binding site sequence which is produced using combinatorial library screening methods to produce a sequence which exhibits high efficiency binding to a preselected iteron sequence, or which has been selected for specific binding to one or a few related iteron sequences.

In another embodiment, the invention contemplates a nucleic acid comprising a geminivirus iteron nucleotide sequence that defines a geminivirus iteron capable of binding to a Rep protein. Specifically, the iteron sequence can be used to compete for binding to geminivirus Rep protein, and thereby prevent Rep protein from binding iteron sequences present on infective geminiviral replicative forms, thereby inhibiting virus replication and preventing symptoms of infection in the plant.

Thus, in one embodiment, the invention contemplates a method for producing resistance to geminivirus infection in a plant comprising introducing a nucleotide sequence into the plant, wherein the nucleotide sequence comprises a geminivirus iteron sequence capable of binding a Rep protein.

Various iteron sequences can be used as described herein, including the use of multiple different iteron sequences to bind multiple Rep proteins from multiple strains of different geminiviruses. In one embodiment, a single nucleic acid molecule may contain multiple iteron sequences. In a preferred embodiment, the nucleic acid comprises each class of iteron sequences shown in Figs1A-1C.

Introduction of a defective Rep protein or iteron nucleotide sequence into plants can be accomplished by a variety of methods including standard gene transfer methods, inoculation of the plant with a transfer or carrier vector, "biolistic" (i.e., ballistic) introduction of nucleic acids into mature plant tissue, direct DNA uptake into plant protoplast, transformation of plants with

Agrobacterium tumefaciens-based vectors, and the like. The Rep protein is typically expressed using a nucleotide sequence which encodes a defective Rep protein and which contains expression control elements which provide for expression of the protein in plants. Alternatively, the DNA can contain the iteron sequence, and no protein expression is required.

Plant expression elements for a nucleotide sequence are generally well known in the art and are not to be considered limiting to the invention. The nucleotide sequence which encodes the Rep protein can be present on an expression vector, as a DNA fragment, or as a component of a "transfer" or carrier vector such as the infectious *Agrobacterium* gene transfer system commonly used in plants.

Thus, iteron sequences can be introduced into plants by a variety of methods, and therefore the invention is not to be construed as so limited. The nucleotide sequence can be introduced directly such as by biolistics, can be present on a vector capable of transcribing a nucleic acid copy of the iteron sequence inside a plant cell, or can be present as a component part of the plant genome of a transgenic plant. Preferred iteron sequences are described in the Examples.

The introduction of a Rep-iteron antagonist, ie., a defective Rep protein, single or multiple iteron nucleotide sequence, according to the invention is effective at inhibiting replication of any geminivirus that infects plants and whose replication depends upon the interaction of functional replication associated protein (Rep) with the iteron sequence of the infecting viral genome. Such antagonist may also include mutated Rep or iterons. Preferred viruses are the *Geminiviridae* family of viruses, which includes *Mastrevirus*, *Curtovirus*, *Topcuvirus* and *Begomovirus* genera, and the *Nanovirus* and *Circoviridae* viruses or other viruses that replicate using Rep proteins that bind to an iteron, wherein use of a Rep-iteron antagonist reduces infection of the organism.

Preferred *Mastrevirus* genus species are selected from the group consisting of Bajra streak virus, Bean yellow dwarf virus, Bromus striate mosaic virus, Chickpea chlorotic dwarf virus, Chloris striate mosaic virus, Digitaria streak virus, Digitaria striate mosaic virus, Maize streak virus//Ethiopia, Maize streak

virus//Ghana1, Maize streak virus//Ghana2, Maize streak virus//Kenya, Maize streak virus//Komatipoort, Maize streak virus//Malawi, Maize streak virus//Mauritius, Maize streak virus//Mozambique, Maize streak virus//Nigeria1, Maize streak virus//Nigeria2, Maize streak virus//Nigeria3, Maize streak virus//Port Elizabeth, Maize streak virus//Reunion1, Maize streak virus//Reunion2, Maize streak virus//Setaria, Maize streak virus//South Africa, Maize streak virus//Tas, Maize streak virus//Uganda, Maize streak virus//Vaalhart maize, Maize streak virus//Vaalhart wheat, Maize streak virus//Wheat-eleusian, Maize streak virus//Zaire, Maize streak virus//Zimbabwe1, Maize streak virus//Zimbabwe2, Miscanthus streak virus, Panicum streak virus/Karino, Panicum streak virus/Kenya, Paspalum striate mosaic virus, Sugarcane streak virus//Egypt, Sugarcane streak virus/Natal, Sugarcane streak virus/Mauritius, Tobacco yellow dwarf virus, Wheat dwarf virus/Czech Republic (Wheat dwarf virus-CJI, WDV-CJI), Wheat dwarf virus/France and Wheat dwarf virus/Sweden.

Preferred *Curtovirus* genus species are selected from the group consisting of Beet curly top virus-California, Beet curly top virus-California//Logan, Beet curly top virus-CFH, Beet curly top virus//Iran, Beet curly top virus-Worland,Horseradish curly top virus, Tomato leafroll virus and Tomato pseudo-curly top virus.

Preferred *Begomovirus* genus species are selected from the group consisting of Abutilon mosaic virus, Acalypha yellow mosaic virus, African cassava mosaic virus//Ghana, African cassava mosaic virus/Kenya, African cassava mosaic virus/Nigeria, African cassava mosaic virus/Uganda, Ageratum yellow vein virus, Althea rosea enation virus, Asystasia golden mosaic virus, Bean calico mosaic virus, Bean dwarf mosaic virus, Bean golden mosaic virus-Brazil, Bean golden mosaic virus-Puerto Rico, Bean golden mosaic virus-Puerto Rico/Dominican Rep.(Bean golden mosaic virus-Dominican Rep., BGMV-DR), Bean golden mosaic virus-Puerto Rico/Guatemala (Bean golden mosaic virus-Guatemala, BGMV-GA], Bhendi yellow vein mosaic virus, Chino del tomate virus (Tomato leaf crumple virus, TLCrV), Cotton leaf crumple virus, Cotton leaf curl virus-India, Cotton leaf curl virus-Pakistan1/Faisalabad1 (Cotton leaf curl

virus-Pakistan2), Cotton leaf curl virus-Pakistan1/Faisalabad2 (Cotton leaf curl virus-Pakistan3),Cotton leaf curl virus-Pakistan1/Multan (Cotton leaf curl virus-Pakistan1), Cotton leaf curl virus-Pakistan2/Faisalabad (Pakistani cotton leaf curl virus), Cowpea golden mosaic virus, Croton yellow vein mosaic virus//Lucknow, Dolichos yellow mosaic virus, East African cassava mosaic virus/Kenya, East African cassava mosaic virus/Malawi, East African cassava mosaic virus/Tanzania, East African cassava mosaic virus/Uganda//1 (African cassava mosaic virus-Uganda variant), East African cassava mosaic virus/Uganda//2, Eclipta yellow vein virus, Eggplant yellow mosaic virus, Eupatorium yellow vein virus, Euphorbia mosaic virus, Honeysuckle yellow vein mosaic virus, Horsegram yellow mosaic virus, Indian cassava mosaic virus, Jatropha mosaic virus, Leonurus mosaic virus, Limabean golden mosaic virus, Lupin leaf curl virus, Macroptilium golden mosaic virus-Jamaica//2, Macroptilium golden mosaic virus-Jamaica//3, Macrotyloma mosaic virus, Malvaceous chlorosis virus, Melon leaf curl virus, Mungbean yellow mosaic virus, Okra leaf curl virus//Ivory Coast, Okra leaf curl virus//India, Papaya leaf curl virus, Pepper huasteco virus, Pepper golden mosaic virus, (Texas pepper virus), Pepper mild tigrÄ virus, Potato yellow mosaic virus//Guadeloupe, Potato yellow mosaic virus/Trinidad and Tobago, Potato yellow mosaic virus/Venezuela, Pseuderanthemum yellow vein virus, Rhynchosia mosaic virus, Serrano golden mosaic virus, Sida golden mosaic virus/Costa Rica, Sida golden mosaic virus/Honduras, Sida golden mosaic virus/Honduras//Yellow vein, Sida yellow vein virus, Solanum apical leaf curl virus, Soybean crinkle leaf virus, Squash leaf curl virus, Squash leaf curl virus/Extended host, Squash leaf curl virus/Restricted host, Squash leaf curl virus/Los Mochis, Squash leaf curl virus-China, Tomato golden mosaic virus/Common strain, Tomato golden mosaic virus/Yellow vein strain, Tobacco leaf curl virus//India, Tobacco leaf curl virus-China, Tomato leaf curl virus//Solanum species D1, Tomato leaf curl virus//Solanum species D2, Tomato leaf curl virus-Australia, Tomato leaf curl virus-Bangalore1 (Indian tomato leaf curl virus-Bangalorel), Tomato leaf curl virus-Bangalore2, (Indian tomato leaf curl virus, ItoLCV), Tomato leaf curl virus-Bangalore3 (Indian tomato

leaf curl virus- BangaloreII), Tomato leaf curl virus-New Delhi/Severe (Tomato leaf curl virus-India2, ToLCV-IN1), Tomato leaf curl virus-New Delhi/Mild (Tomato leaf curl virus-India2, ToLCV-IN2), Tomato leaf curl virus-New Delhi/Lucknow (Indian tomato leaf curl virus), Tomato leaf curl virus//Senegal, Tomato leaf curl virus-Sinaloa (Sinaloa tomato leaf curl virus, STLCV), Tomato leaf curl virus-Taiwan, Tomato leaf curl virus-Tanzania, Tomato mottle virus, Tomato mottle virus-Taino (Taino tomato mottle virus, TTMoV), Tomato severe leaf curl virus//Guatemala, Tomato severe leaf curl virus//Honduras, Tomato severe leaf curl virus//Nicaragua, Tomato yellow dwarf virus, Tomato yellow leaf curl virus-China, Tomato yellow leaf curl virus-Israel, Tomato yellow leaf curl virus-Israel/Mild, Tomato yellow leaf curl virus-Israel/Egypt, (Tomato yellow leaf curl virus-Egypt, TYLCV-EG), Tomato yellow leaf curl virus-Israel//Cuba, Tomato yellow leaf curl virus-Israel//Jamaica, Tomato yellow leaf curl virus-Israel//Saudi Arabia1, (Tomato yellow leaf curl virus-Northern Saudi Arabia, TYLCV-NSA), Tomato yellow leaf curl virus-Nigeria, Tomato yellow leaf curl virus-Sardinia, Tomato yellow leaf curl, virus-Sardinia/Sicily (Tomato yellow leaf curl virus-Sicily, TYLCV-SY), Tomato yellow leaf curl virus-Sardinia/Spain//1 (Tomato yellow leaf curl virus-Spain, TYLCV-Sp), Tomato yellow leaf curl virus-Sardinia/Spain//2 (Tomato yellow leaf curl virus-Almeria, TYLCV-Almeria ) Tomato yellow leaf curl virus-Sardinia/Spain//3 (Tomato yellow leaf curl virus-European strain), Tomato yellow leaf curl virus-Saudi Arabia (Tomato yellow leaf curl virus-Southern Saudi Arabia, TYLCV-SSA), Tomato yellow leaf curl virus-Tanzania, Tomato yellow leaf curl virus-Thailand//1, Tomato yellow leaf curl virus-Thailand//2 , Tomato yellow leaf curl virus//Yemen, Tomato yellow mosaic virus-Brazil//1, Tomato yellow mosaic virus-Brazil//2, Tomato yellow mottle virus, Tomato yellow vein streak virus-Brazil, Watermelon chlorotic stunt virus, Watermelon curly mottle virus and Wissadula golden mosaic virus-Jamaica//1.

The invention also contemplates a nucleic acid molecule, such as a DNA expression vector, useful for expression of a Rep protein of this invention in plants. Thus the nucleic acid molecule contains a nucleotide sequence which

encodes a geminivirus Rep protein, variant or fragment thereof capable of binding a geminivirus iteron nucleotide sequence, and further contains elements for regulation and control of gene expression in plants. Exemplary elements are described in United States Patent Nos. 5,188,642, 5,202,422, 5,463,175 and 5,639,947, the disclosures of which are hereby incorporated by reference.

Exemplary expression vectors and systems for introduction of a Rep protein into plants are described in the Examples.

The invention further contemplates a transgenic plant containing a nucleotide sequence of this invention for expressing the geminivirus Rep protein, variants and fragments thereof. Preparation of transgenic plants is well known in the art and described at least in the above-mentioned U.S. patents.

Also contemplated is a composition useful for introducing a nucleotide sequence of this invention into plants. The composition comprises an effective amount of the nucleotide sequence for introducing the geminivirus Rep protein into a plant, and depends upon the method used for introducing the protein to the plant. For example, using direct DNA uptake by protoplast, the composition is an aqueous solution containing nucleic acid and buffers to facilitate uptake by protoplast, as is well known. For transformation by *Agrobacterium*, the composition contains a suspension of *Agrobacteria* containing the nucleotide sequence capable of expressing the dsDNA-binding protein.

In one embodiment, a vector is employed that is capable of integrating the desired gene sequences into the host cell chromosome. Cells that have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker may provide for prototrophy to an auxotrophic host, biocide resistance, e.g., antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection.

In another embodiment, the introduced sequence will be incorporated into a plasmid or viral vector capable of autonomous replication in the recipient

host. Any of a wide variety of vectors may be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector may be recognized and selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

DNA encoding the desired protein is preferably operably linked to a promoter region, a transcription initiation site, and a transcription termination sequence, functional in plants. Any of a number of promoters which direct transcription in a plant cell is suitable. The promoter can be either constitutive or inducible. Some examples of promoters functional in plants include the nopaline synthase promoter and other promoters derived from native Ti plasmids, viral promoters including the 35S and 19S RNA promoters of cauliflower mosaic virus (Odell *et al., Nature 313:*810-812 (1985)), and numerous plant promoters.

Alternative promoters that may be used include nos, ocs, and CaMV promoters. Overproducing plant promoters may also be used. Such promoters, operably linked to the Rep gene, should increase the expression of the Rep protein. Overproducing plant promoters that may be used in this invention include the promoter of the small subunit (ss) of ribulose-1,5-biphosphate carboxylase from soybean (Berry-Lowe *et al., J. Molecular and App. Gen. 1*:483-498 (1982), and the promoter of the chlorophyll a/b binding protein. These two promoters are known to be light-induced in eukaryotic plant cells (see, for example, *Genetic Engineering of Plants, an Agricultural Perspective*, A. Cashmore, Plenum, New York 1983, pages 29-38; Corruzi, G. *et al., J. of Biol. Chem. 258*:1399 (1983); and Dunsmuir, P. *et al., J. of Mol. and Applied Genet. 2*:285 (1983)).

Genetic sequences comprising the desired gene or antisense sequence operably linked to a plant promoter may be joined to secretion signal sequences and the construct ligated into a suitable cloning vector. In general, plasmid or

viral (bacteriophage) vectors containing replication and control sequences derived from species compatible with the host cell are used. The cloning vector may typically carry a replication origin, as well as specific genes that are capable of providing phenotypic selection markers in transformed host cells, typically antibiotic resistance genes.

General methods for selecting transgenic plant cells containing a selectable marker are well known and taught, for example, by Herrera-Estrella, L. and Simpson, J. (1988) "Foreign Gene Expression in Plants" in *Plant Molecular Biology, A Practical Approach*, Ed. C.H. Shaw, IRL Press, Oxford, England, pp. 131-160, Scholthof, H. *J. Virol 73*:7823-7829 (1999), Yusibov *et al., Curr. Top. Microbiol. Immunol. 240*81-94 (1999), Houdred *et al., Plant Physiol. 119*:713-724 (1999).

In another embodiment, the present invention relates to a transformed plant cell comprising exogenous copies of DNA (that is, copies that originated outside of the plant) encoding a Rep gene expressible in the plant cell wherein said plant cell is free of other foreign marker genes (preferably, other foreign selectable marker genes); a plant regenerated from the plant cell; progeny or a propagule of the plant; and seed produced by the progeny.

Plant transformation techniques are well known in the art and include direct transformation (which includes, but is not limited to: microinjection (Crossway, *Mol. Gen. Genetics 202*:179-185 (1985)), polyethylene glycol transformation (Krens *et al., Nature 296*:72-74 (1982)), high velocity ballistic penetration (Klein *et al., Nature 327*:70-73 (1987)), fusion of protoplasts with other entities, either minicells, cells, lysosomes, or other fusible lipid-surfaced bodies (Fraley *et al., Proc. Natl. Acad. Sci. USA 79*:1859-1863 (1982)), electroporation (Fromm *et al., Proc. Natl. Acad. Sci. USA 82*:5824 (1985)) and techniques set forth in U.S. Patent No. 5,231,019)) and *Agrobacterium tumefaciens* mediated transformation as described herein and in (Hoekema *et al., Nature 303*:179 (1983), de Framond *et al., Bio/technology 1*:262 (1983), Fraley *et al.* WO84/02913, WO84/02919 and WO84/02920, Zambryski *et al.* EP 116,718, Jordan *et al., Plant Cell Reports* 7:281-284 (1988), Leple *et al.*

*Plant Cell Reports 11:*137-141 (1992), Stomp *et al., Plant Physiol. 92:*1226-1232 (1990), and Knauf *et al., Plasmid 8:*45-54 (1982)). Another method of transformation is the leaf disc transformation technique as described by Horsch *et al. Science 227:*1229-1230 (1985).

The transformation techniques can, for example, utilize a DNA encoding an amino acid sequence of Figs.1A-1C, fragments thereof or the antisense sequence, expressible in plants. Included within the scope of a gene encoding the Rep sequences of Figs.1A-1C are functional derivatives, as well as variant, analog, species, allelic and mutational derivatives of the polypeptides of the invention.

As used herein, modulation of Rep expression may entail the enhancement or reduction of the naturally occurring levels of the protein. Specifically, the translation of RNA encoding Rep may also be reduced or inhibited by the expression of an antisense gene or RNA.

In general, antisense cloning entails the generation of an expression module which encodes an RNA complementary (antisense) to the RNA encoding Rep (sense). By expressing the antisense RNA in a cell which expresses the sense strand, hybridization between the two RNA species will occur resulting in the blocking of translation. Alternatively, overexpression of the Rep protein might be accomplished by use of appropriate promoters, enhancers, and other modifications. Those of skill in the art would be aware of references describing the use of antisense genes in plants (van der Krol *et al., Gene 72:*45-50 (1988); van der Krol *et al., Plant Mol. Biol. 14:*467-486 (1990); Zhang *et al., Plant Cell 4:*1575-1588 (1992) and U.S. Pat. 5,107,065).

Other foreign marker genes (i.e., exogenously introduced genes) typically used include selectable markers such as a *neo* gene (Potrykus *et al., Mol. Gen. Genet 199:*183-188 (1985)) which codes for kanamycin resistance; a *bar* gene which codes for bialaphos resistance; a mutant EPSP synthase gene (Hinchee *et al., Bio/technology 6:*915-922 (1988)) which encodes glyphosate resistance; a nitrilase gene which confers resistance to bromoxynil (Stalker *et al., J. Biol. Chem. 263:*6310-6314 (1988)); a mutant acetolactate synthase gene

(ALS) which confers imidazolinone or sulphonylurea resistance (EP application number 154,204); a methotrexate resistant *DHFR* gene (Thillet *et al., J. Biol. Chem. 263:*12500-12508) and screenable markers which include ß-glucuronidase (*GUS*) or an R-locus gene, alone or in combination with a C-locus gene (Ludwig *et al., Proc. Natl. Acad. Sci. USA 86:*7092 (1989); Paz-Ares *et al., EMBO J. 6:*3553 (1987)).

The genetic construct for expressing the desired protein can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. The genetic material may also be transferred into plant cells using polyethylene glycol to form a precipitation complex with the genetic material that is taken up by cells. (Paszkowski *et al., EMBO J. 3:*2717-22 (1984)). The desired gene may also be introduced into plant cells by electroporation. (Fromm *et al.*, "Expression of Genes Transferred into Monocot and Dicot Plant Cells by Electroporation," *Proc. Nat'l. Acad. Sci. U.S.A. 82:*5824 (1985)). In this technique, plant protoplasts are electroporated with plasmids containing the desired genetic construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of plasmids. Electroporated plant protoplasts reform cell walls, divide, and form plant calli. Selection of the transformed plant cells expressing the desired gene can be accomplished using phenotypic markers as described above. Alternatively, microprojectile bombardment may be used (Daniel H. *Methods Mol. Biol. 62:*463-489 (1997).

Another method of introducing the desired gene into plant cells is to infect the plant cells with *Agrobacterium tumefaciens* transformed with the desired gene. Under appropriate conditions well-known in the art, transformed plant cells are grown to form shoots, roots, and develop further into plants. The desired genetic sequences can be joined to the Ti plasmid of *Agrobacterium tumefaciens*. The Ti plasmid is transmitted to plant cells on infection by *Agrobacterium tumefaciens* and is stably integrated into the plant genome. *See* Horsch *et al.*, "Inheritance of Functional Foreign Genes in Plants," *Science 233*: 496-498 (1984); Fraley *et al., Proc. Nat'l Acad. Sci. U.S.A. 80*: 4803

(1983)); Feldmann, K.A. *et al.*, *Mol. Gen. Genet.*, *208*: 1-9 (1987); Walden, R. *et al.*, *Plant J.*, *1*: 281-288 (1991).

Presently there are several different ways to transform plant cells with *Agrobacterium*:

(1) co-cultivation of *Agrobacterium* with cultured, isolated protoplasts, or

(2) transformation of cells or tissues with *Agrobacterium*.

Method (1) requires an established culture system that allows culturing protoplasts and plant regeneration from cultured protoplasts. Method (2) requires that the plant cells or tissues can be transformed by *Agrobacterium* and that the transformed cells or tissues can be induced to regenerate into whole plants. In the binary system, to have infection, two plasmids are needed: a T-DNA containing plasmid and a *vir* plasmid.

Routinely, however, one of the simplest methods of plant transformation is explant inoculation, which involves incubation of sectioned tissue with *Agrobacterium* containing the appropriate transformation vector (Plant Genetic Transformation and Gene Expression, A Laboratory Manual, Oxford: Blackwell Scientific Publications (1988); Walden, Genetic Transformation in Plants, Milton Koynes: Open University Press (1988)).

Methods for inserting viral DNA into plant material are known in the art, *see* for example, U.S. Patent No. 5,569,597 or Porta C. *et al.*, *5*:209-221 (1996).

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be used for the expression of the desired gene. Suitable plants include, for example but are not limited to, species from the genera *Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manicot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Lycopersion, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Cichorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hemerocallis, Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browallia,*

*Glycine, Lolium, Zea, Triticum, Sorghum*, and *Datura*. Additional plant genera that may be transformed by *Agrobacterium* include *Ipomoea, Passiflora, Cyclamen, Malus, Prunus, Rosa, Rubus, Populus, Santalum, Allium, Lilium, Narcissus, Ananas, Arachis, Phaseolus*, and *Pisum*.

Plant regeneration techniques are well known in the art and include those set forth in the *Handbook of Plant Cell Culture*, Volumes 1-3, Eds. Evans *et al.* Macmillan Publishing Co., New York, NY (1983, 1984, 1984, respectively); Predieri and Malavasi, *Plant Cell, Tissue, and Organ Culture 17*:133-142 (1989); James, D.J., *et al.*, *J. Plant Physiol. 132*:148-154 (1988); Fasolo, F., *et al.*, *Plant Cell, Tissue, and Organ Culture 16*:75-87 (1989); Valobra and James, *Plant Cell, Tissue, and Organ Culture 21*:51-54 (1990); Srivastava, P.S., *et al.*, *Plant Science 42*:209-214 (1985); Rowland and Ogden, *Hort. Science 27*:1127-1129 (1992); Park and Son, *Plant Cell, Tissue, and Organ Culture 15*:95-105 (1988); Noh and Minocha, *Plant Cell Reports 5*:464-467 (1986); Brand and Lineberger, *Plant Science 57*:173-179 (1988); Bozhkov, P.V., *et al.*, *Plant Cell Reports 11*:386-389 (1992); Kvaalen and von Arnold, *Plant Cell, Tissue, and Organ Culture 27*:49-57 (1991); Tremblay and Tremblay, *Plant Cell, Tissue, and Organ Culture 27*:95-103 (1991); Gupta and Pullman, U.S. Patent No. 5,036,007; Michler and Bauer, *Plant Science 77*:111-118 (1991); Wetzstein, H.Y., *et al.*, *Plant Science 64*:193-201 (1989); McGranahan, G.H., *et al.*, *Bio/Technology 6*:800-804 (1988); Gingas, V.M., *Hort. Science 26*:1217-1218 (1991); Chalupa, V., *Plant Cell Reports 9*:398-401 (1990); Gingas and Lineberger, *Plant Cell, Tissue, and Organ Culture 17*:191-203 (1989); Bureno, M.A., *et al.*, *Phys. Plant. 85*:30-34 (1992); and Roberts, D.R., *et al.*, *Can. J. Bot. 68*:1086-1090 (1990).

Plant regeneration from cultured protoplasts is described in Evans *et al.*, "Protoplast Isolation and Culture," in *Handbook of Plant Cell Culture 1*:124-176 (MacMillan Publishing Co., New York, 1983); M.R. Davey, "Recent Developments in the Culture and Regeneration of Plant Protoplasts," *Protoplasts, 1983 - Lecture Proceedings*, pp. 19-29 (Birkhauser, Basel, 1983); P.J. Dale, "Protoplast Culture and Plant Regeneration of Cereals and Other

Recalcitrant Crops," in *Protoplasts 1983 - Lecture Proceedings*, pp. 31-41 (Birkhauser, Basel, 1983); and H. Binding, "Regeneration of Plants," in *Plant Protoplasts*, pp. 21-37 (CRC Press, Boca Raton, 1985).

Techniques for the regeneration of plants varies from species to species but generally, a suspension of transformed protoplasts containing multiple copies of the desired gene is first provided. Embryo formation can then be induced from the protoplast suspensions, to the stage of ripening and germination as natural embryos. The culture media will generally contain various amino acids and hormones, such as auxins and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa.

Mature plants, grown from transformed plant cells, are selfed to produce an inbred plant. The inbred plant produces seed containing the recombinant DNA sequences promoting increased resistance to geminivirus infection.

Parts obtained from regenerated plants, such as flowers, seeds, leaves, branches, fruit, and the like are covered by the invention provided that these parts comprise the geminivirus resistant cells. Progeny and variants, and mutants of the regenerated plants are also included within the scope of this invention. As used herein, mutant describes variation as a result of environmental conditions, such as radiation, or as a result of genetic variation in which a trait is transmitted meiotically according to well-established laws of inheritance.

All plants which can be transformed are intended to be hosts included within the scope of the invention (preferably, dicotyledonous plants). Such plants include, but are not limited to, species from the genera *Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Lycopersicon, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Cichorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hererocallis, Nemesia, Pelargonium, Panicum,*

*Pennisetum, Ranunculus, Sencia, Salpiglossis, Cucumis, Browalia, Glycine, Lolium, Zea, Triticum, Sorghum, Malus, Apium, Datura* and *Monocotyledonous* plants such as corn.

Some examples of commercially useful agricultural plants to which methods of the invention may be applied include Abutilon, Acalypha, apple, Ageratum, Althea rosea, Asystasia, Bajra, banana, barley, beans, beet, Blackgram, Bromus, Cassava, chickpea, Chilllies, Chloris, clover, coconut, coffee, cotton, cowpea, Croton, cucumber, Digitaria, Dolichos, eggplant, Eupatorium, Euphorbia, fababean, honeysuckle, horsegram, Jatropha, Leonurus, limabean, Lupin, Macroptilium, Macrotyloma, maize, melon, millet, mungbean, oat, okra, Panicum, papaya, Paspalum, peanut, pea, pepper, pigeon pea, pineapple, Phaseolus, potato, Pseuderanthemum, pumpkin, Rhynchosia, rice, Serrano, Sida, sorghum, soybean, squash, sugarcane, sugarbeet, sunflower, sweet potato, tea, tomato, tobacco, watermelon, wheat and Wissadula.

Additionally, the method may be applied to grains, legumes, vegetables and fruits, including but not limited to wheat, corn, barley, alfalfa, cotton, rapeseed, rye, peas, celery, grapes, cabbage, oilseed, apples, strawberries, mulberries, cranberries and lettuce that become infected with geminivirus.

Other embodiments of the invention will be apparent to one skilled in the art in light of the previous disclosures. Therefore, the previous disclosures as well as the following examples should not be considered to limit the scope of the invention.

### *Example 1*
### *Characteristics of Replication Associated Protein (Rep)*

Initiation of replication is a function of site specific binding of the Rep protein to its cognate site in the common region. In most of the cases studied so far, the binding site comprises of a small sequence of 5-8 bases in the origin of replication that are repeated and occur in close proximity to the TATA box. Each of these repeats is referred to as iterons. For ToLCV-Nde the binding site of the

Rep protein has not been determined so far. Electromobility shift assays (EMSA) were used to identify the Rep protein binding site in the intergenic region of the viral genome and determine the specificity of this interaction using two homologous strains of ToLCV-Nde which differ in their binding site sequence in the common region. Transient expression in tobacco protoplasts indicated that the Rep proteins of the two strains are not interchangeable despite 91% sequence identity between the two genes. In addition, this specificity of replication was found associated with the N-terminal sequences of the Rep protein that in turn may influence its interaction with the origin of replication.

***Identification of the binding site of replicase associated protein in the intergenic region of mild and severe strain of ToLCV-Nde***

Based upon analogy with published reports potential repeat motifs were identified in the common region of mild and severe strain of ToLCV-Nde close to the TATA (SEQ ID NO:158) box. The ability of the Rep protein to interact with this repeat motif in electrophoretic mobility shift assays (EMSAs) was determined. The binding site sequence in the severe strain was identified as GGTGTCTGGAGTC (SEQ ID NO:121 ) while in the mild strain the repeat motif was GGCGTCTGGCGTC GGTGTCTGGAGTC (SEQ ID NO:159). A 13 nucleotide sequence was used comprising this repeat in binding assays. As a control, the 52 bp common region was used as a probe to interact with the Rep protein. Results showed the formation of a similar complex with both, the identified 13 mer sequence as well as the full length common region suggesting that the identified sequence may well be the binding site for the Rep protein. Similarly, the Rep protein of the mild strain could shift either of the probes resulting in complex. The specificity of this binding was confirmed by competition assays. For both the mild and the severe strain, the complex was abolished when an excess of cold 13-mer probe specific to each strain was added as a specific competitor in the binding reaction. Further, a 1000 fold excess of an unrelated 18-mer oligonucleotide did not affect the formation of complex in either of the strains.

***Binding of the rep protein to its cognate site is specific between the strains***

To determine if the Rep protein of the mild strain can distinguish its cognate binding site from the severe strain, the 19 mer repeat sequence of the severe strain was used as a probe. In gel shifts, the Rep protein of the mild strain did not form any complex with the severe strain binding site. In analogous experiments with the Rep protein of the severe strain, a very weak complex was observed when the binding site sequence of the mild strain was used as a probe. These results indicated that the Rep proteins of the two strains bind specifically to their cognate sites in the origin of replication.

***Specificity of binding by the Rep protein may depend on several factors***

Earlier results suggested that binding of the Rep protein to its cognate site may be a specific event which prompted us to investigate the factors influencing the specificity of binding. Synthetic oligonucleotides were designed to address several parameters. These were a the sequence of the repeat motif, i.e. changing the 5′ or the 3′ iteron from the related strain or any unrelated virus sequence, b) the spacing within the two repeat motifs and c) the numbers of the repeat motifs: one, two or four.

Sequence of the repeat motif:

5′ severe, 3′ mild, (GGTGTCTGGCGTC) (SEQ ID NO:160): The Rep protein of the mild strain did not show any complex formation with this probe while the severe strain Rep protein formed a complex as visualized by a shift in the probe.

5′ unrelated, 3′ mild, (GGAGTCTGGCGTC)(SEQ ID NO:161): Neither of the Rep proteins from the mild or the severe strain could generate a complex with this probe.

5′ unrelated , 3′ severe: (GGAGTCTGGTGTC)(SEQ ID NO:162): The mild strain Rep protein did not form any complex with this probe but the severe strain formed a very weak complex.

5′ mild , 3′ unrelated: (GGCGTCTGGAGTC)(SEQ ID NO:163): Very weak complex formation was observed by the Rep proteins in both strains.

Results of the study showed no complex formation was observed if the spacing between the two repeats was either increased or decreased from the 3 bases already known to exist. Additionally, results with the binding assays showed that the Rep protein of the severe strain was unable to bind efficiently if only a single iteron (GGTGTC) (SEQ ID NO:164) was presented as a probe, but with two iterons (GGTGTC T GGTGTC) (SEQ ID NO:165) the binding was more specific and a retarded complex was clearly distinguishable. When the number of repeat motifs was doubled (GGTGTC T GGTGTC T GGTGTCT GGTGTC) (SEQ ID NO:166), at least three different bands with retarded mobility were observed, suggesting that the Rep protein recognizes and binds to additional sites.

Doubling the number of available binding sites generated multiple bands.

***Binding domain of Rep protein may lie on its N-terminus***

In order to define the minimal size of Rep protein responsible for binding to the origin, three different truncated Rep proteins from the severe strain were purified. T-Rep1 has only the 1-52 amino acids of the Rep protein and consists of the motif FLTYPKC (SEQ ID NO:172), a conserved sequence present in all the organisms that replicate via a rolling circle mechanism and the helices α-1 and α-2. T-Rep2 has the 1-111 amino acids from the N-terminus of the Rep protein and comprises of all the three conserved motifs 1, 2 and 3 as well as the helices α-1 and α-2. T-Rep3 has the amino acids 1-160 from the N-terminus of the Rep. All three proteins were checked in gel shift assays using the 19 mer repeat sequence to confirm if they still retain DNA binding activity. Of the three truncated Rep proteins, T-Rep 1 and 2 formed a weak complex that appeared as a faint, retarded band, but T-Rep3 generated a complex similar to the wild type full length protein, suggesting that amino acids 1-160 may contain sufficient information to allow *in vitro* binding of the Rep protein to its cognate site in the origin of replication.

To confirm if the observed bands with retarded mobility were specifically due to binding of the truncated Rep proteins to the 19 mer binding site repeat

motif, a supershift assay was performed using the α- Histidine antibody. (Ausubel *et al. Protocols in Molecular Biology*) The purified full length and truncated Rep proteins were allowed to incubate with the α-His antibody on ice followed by incubation with the repeat motif as the probe. In all the cases, a supershift was observed suggesting that the interaction of the Rep proteins with the binding site was specific. Without the α-His-antibody, the complex formed in all cases migrated much faster on the gel indicating that the supershift was specifically due to the binding of the truncated Rep proteins to its iteron sequence as well as to the antibody. In all the cases mentioned, the shifts could be abolished by using excess of specific competitor DNA.

In another experiment, the full length and truncated rep protein were immunoprecipitated using AC1 antibody . All the three truncated Rep proteins could be immunoprecipitated with anti His antibody showing that the complex formation and subsequent trapping was due to specific interaction of the Rep protein with its antibody.

Southwestern assays were used to confirm the specificity of the shifts observed with the truncated Rep proteins, a southwestern blot was undertaken. Full length Rep protein as well as the truncated versions of it were run on an SDS PAGE and transferred on a nitrocellulose membrane The transferred proteins were then detected using the 13 mer repeat motif as the probe. All the three truncated Reps could be detected in the southwestern indicating that only 1-56 amino acids of the Rep protein may be enough to bind or recognize the binding site in the origin of replication.

### *Example 2*
### *Sequence Parameters that Determine Rep Binding Specificity*

To determine the parameters that may influence specificity of binding of the Rep protein to its cognate site in the *ori*, two homologous strains of tomato leaf curl virus from New Delhi (ToLC-Nde ) were used. These strains share 94% sequence identity but cause very different symptoms on tomato and *Nicotiana benthamiana* plants (Padidam, M., *et al., J. Gen. Virol. 76*:25-35 (1995)). While

the severe strain is characterized by severe puckering and downward leaf curling in the plants, the mild strain produces mild symptoms with minor leaf curl and no puckering of the leaves. In addition, the two strains do not support the replication of the heterologous DNA (Chatterji, A., *et al., J. Virol 73*:5541-5549 (1999)) making them ideal experimental systems to investigate specificity of interaction between the Rep protein and its binding site in the origin of replication.

Previous studies (Chatterji, A., *et al., J. Virol 73*:5541-5549 (1999)) indicated a possible interaction between amino acid 10, near the N-terminus of the Rep protein and the putative binding site in the origin that may determine specificity of replication between the severe and the mild strains of ToLCV-Nde. Exchange of the 10th amino acid between the Rep proteins of the two strains of ToLC-NdeV with a concomitant change in the binding site sequence in the origin altered the replication of the two strains suggesting that these components may influence the levels of viral replication and accumulation (Chatterji, A., *et al., J. Virol 73*:5541-5549 (1999)).

An objective of this study was to determine if the repeat sequences identified earlier in the origin of replication of the two strains function as the binding sites for their respective Rep proteins. Further, it was of interest to define DNA sequence requirements for specificity of origin recognition in the Rep proteins of the two strains by using chimeric iteron sequences.

Electrophoretic mobility shift assays (EMSAs) were performed using different synthetic oligonucleotides as probes or competitors to show specificity of binding in our assays. The nature and significance of DNA-protein interaction was studied *in vivo* using transient replication assays in tobacco protoplasts. Alterations were found with respect to sequence, size or number of iterons that reduced binding by the Rep protein and resulted in drastic reduction in virus replication. In addition, it was found that the inability of the Rep protein of the mild strain to accumulate DNA B of the severe strain was related to its inability to recognize the binding site sequence of the severe strain DNA-B.

*Materials and Methods*

A.. Expression of ToLCV Rep proteins

The full length AC1 gene from the severe and the mild strains of ToLC-NdeV was amplified from pMPA1 (DNA-A of the severe strain, ToLC-NdeV) and pMPA2 (DNA-A of the mild strain, ToLC-NdeV) (Padidam, M., *et al., J. Gen. Virol. 76*:25-35 (1995)). The amplified sequence was ligated between *Bam* HI and *Hind* III sites in baculovirus expression vector, pBAC4x-1 (Novagen) resulting in an in-frame fusion of the AC1 gene sequence with the vector sequence encoding a methionine and six histidine residues under the *polh* promoter. The clones were identified and confirmed by restriction digestion and sequence analysis.

Recombinant baculovirus was isolated by co-transfecting 0.5 (g of recombinant plasmid with 1 g of linearized *Autographa californica* nuclear polyhedrosis virus DNA (Smith, G.E., and Summers, M.D., *Virology 89*:517-527 (1978)) into *Spodoptera frugiperda* Sf9 cells (Summers, M.D., and Smith, G.E., *Texas Agricultural Experiment Station Bulletin No. 1555* (1987)). Recombinant viruses were plaque-purified and a high titer stock was prepared. Large-scale purification of the target protein was done in *Trichoplusia ni* Tn-5 cells (Invitrogen).

B. Purification of Rep proteins

Tn-5 cells were harvested 60 h post infection by centrifugation at 3000 rpm for 10 minutes. The pellets were washed in 1X PBS and suspended in ice cold 1X binding buffer (5 mM imidazole, 0.5 M NaCl, and 20 mM Tris, pH 7.9). The cells were lysed by three cycles of freeze-thaw and the lysate was clarified at 15,000 rpm for 30 minutes. The resulting supernatant was loaded on a Ni-NTA column (Novagen) previously equilibrated with binding buffer and washed with 10 column volumes of wash buffer (70 mM imidazole, 0.5 mM NaCl, and 20 mM Tris, pH7.9). The protein was eluted with 1M imidazole, 0.5 mM NaCl and 20 mM Tris, pH7.9. The eluted fractions were dialyzed against 20mM Tris, pH7.9,

150mM NaCl to remove imidazole, concentrated using Centricon filters (Amicon) and protein concentration was estimated using Bradford's reagent (Biorad).

C. Electrophoretic mobility shift assays

ToLC-NdeV specific primers were used to amplify a 52 bp fragment from the IR of the virus genome. This fragment contains the iterons, the transcription start site as well as the TATA box and the conserved hairpin sequence. The amplified fragment was end labeled with ($^{32}$P ATP and T4 polynucleotide kinase, purified on polyacrylamide gels and was used as a probe in the EMSAs. The 18mer oligonucleotides containing the potential binding sites (underlined) for the Rep proteins of the two strains were synthesized and annealed to their complementary strands. These two oligonucleotide probes were named bs-m, 5'-GGCGTCTGGCGTCT-3' (U15017) (SEQ ID NO:180) for the mild strain and bs-s, 5'- GGTGTCTGGAGTCT-3' (U15015) (SEQ ID NO:189) for the severe strain. The final concentrations of the probes were 500 pM (30,000cpm). The concentration of competitor DNA used was 100 pM per reaction. Probe and competitor DNAs were purified on Sephadex G-25 columns, quantified by a scintillation counting followed by dilution to 30,000 cpm for the binding assays.

The binding assays (EMSAs) were performed using the purified Rep protein from the two strains. The binding reactions contained 500 ng of pure protein, 1 ng of labeled DNA and 0.2 $\mu$g of poly dI-dC. Binding buffer contained 20mM HEPES pH 7.5, 60 mM KCl, 1mM DTT and 15 % glycerol. Reactions were incubated at 25$^{O}$ C for 30 minutes and the complexes were resolved on 4% polyacrylamide gels in 0.25X TBE. The gels were dried on Whatman paper and autoradiographed. Comparative efficiency of binding was analyzed by quantifying the amount of radioactivity in the retarded bands using the phosphorimager (Molecular Dynamics).

The sequences of the synthetic oligonucleotides used as probes or competitors in the EMSAs are listed below in Table 1.

**TABLE 1**
**Comparative levels of *in vitro* binding and *in vivo* replication by the origin mutants**

| Mutant | Oligonucleotide sequence | Binding[a,c] | Replication[b] | |
|---|---|---|---|---|
| | | | ss[d] | sc[d] |
| CR-s | GGTGTCTGGAGTC (SEQ ID NO: 121) | 100 | 100 | 100 |
| CR-m | GGCGTCTGGCGTC (SEQ ID NO: 123) | 100 | 100 | 89.53 |
| IT-1/2(s) | GGTGTCTGGCGTC (SEQ ID NO: 168) | 100 | 100 | 93.41 |
| IT-1/2(m) | GGTGTCTGGCGTC (SEQ ID NO: 168) | <1 | <1 | <1 |
| IT-3/4(s) | GGGGTCTGGAGTC (SEQ ID NO:51) | 13.4 | 28.61 | 2.18 |
| IT-3/4(m) | GGGGTCTGGAGTC (SEQ ID NO: 52) | <1 | <1 | <1 |
| IT-5/6(s) | GGGGTCTGGCGTC (SEQ ID NO: 170) | <1 | 3.23 | <1 |
| IT-5/6(m) | GGGGTCTGGCGTC (SEQ ID NO: 169) | <1 | <1 | <1 |
| IT-7/8(s) | GGCGTCTGGGGTC (SEQ ID NO: 171) | <1 | 2.84 | <1 |
| IT-7/8(m) | GGCGTCTGGGGTC (SEQ ID NO: 171) | 18.41 | <1 | <1 |
| IT-9/10(S) | GGTGTCTGGTGTC (SEQ ID NO: 53) | 102.6 | 108 | 103.9 |
| IT-11/12(S) | GGAGTCTGGAGTC (SEQ ID NO:58) | 87.21 | 98.63 | 98.82 |
| IT-13/14(S) | GGTGTCTTTTTTGGAGTC | <1 | 4.47 | <1 |

| | | | | | |
|---|---|---|---|---|---|
| | | (SEQ ID NO:63) | | | |
| | IT-13/14(M) | GGCGTCTTTTTTGGCGTC (SEQ ID NO:192) | >1 | <1 | <1 |
| | IT-15/16(S) | GGTGTCGGAGTC (SEQ ID NO: 174) | >1 | <1 | <1 |
| | IT-15/16(M) | GGCGTCGGCGTC ((SEQ ID NO: 175) | >1 | <1 | 1.53 |
| | IT-17/18(S) | GGTGTC (SEQ ID NO: 164) | 29.47 | 24.63 | <1 |
| | IT-17/18(M) | GGCGTC (SEQ ID NO: 176) | 8.53 | 17.64 | 0.13 |
| | IT-19/20(S) | GGTGTCTGGTGTCTGGTG-TCTGGTGTC (SEQ ID NO:191) | | 75.94 | 47.13 |
| | IT-19/20(M) | GGCGTCTGGCGTCTGGCGTCTGGCGTC (SEQ ID NO: ) | | - 62.24 | 36.53 |

[a]The values shown represent the amount of radioactivity (%) bound in the shifted DNA-protein complex band as a result of the Rep protein binding to the $^{32}P$ labeled DNA protein in gel shift assays.

[b]The values shown are average (%) amounts of single stranded (ss) and supercoiled (sc) viral DNA detected in four independent protoplasts transfections per mutant. Protoplasts prepared from *N. tabacum* BY2 cells were transfected with 2 $\mu$g of DNA-A and harvested 48h after electroporation. Viral DNA was quantitated using a phosphorimager. Standard error values between different transfections were in the range of $\pm$ 2-5%.

[c]The amount of radioactivity bound in the complex shifted as a result of Rep protein binding to its respective CR sequences was assigned a value of 100.

[d]The amounts of viral DNA observed in protoplasts inoculated with the wild type DNA-A of the severe or the mild strain were assigned a value of 100.

D. Immunoprecipitation assays

Rep protein was immunopreciptated from 50 (µg of Sf9 cell extract using 10 (µg of anti-AC1 antiserum and rabbit anti-mouse IgG coupled to Sepharose beads. The immuno complexes were resolved by SDS-PAGE and detected by immunoblotting (Towbin, H., *et al., Proc. Natl. Acad. Sci. USA 76*:4350-4354 (1979)) using a rabbit polyclonal anti-AC1 antiserum and an anti-rabbit goat

antibody conjugated with horseradish peroxidase. The peroxidase activity was detected using chemiluminescent Supersignal substrate (Pierce).

E. Construction of mutants

A 146 bp common region fragment from *Nco* I to *Ssp* I from both severe and mild strains of ToLC–NdeV was amplified and cloned in pBS (SK-) vector. Site directed mutations were made in the iterated sequences using overlapping PCR (Horton, R.M., "In vitro recombination and mutagenesis of DNA", *Methods in Molecular Biology,* PCR cloning protocols, B.A. White, ed., Humana Press. Inc., Totowa, New Jersey (1994), vol. 67, pp.141-150). Following confirmation of mutations by sequencing, the common region fragments were recloned into the DNA-A of the respective strains. For convenience, the mutants were given names identical to the synthetic oligonucleotides used to create the nucleotide changes in the iteron sequence.

F. Transient Replication assays

Protoplasts isolated from *Nicotiana tabacum* BY2 cells were electroporated and cultured according to published methods (Watanabe, Y., *et al., FEBS Lett. 219*:65-69 (1987)). Transfections were done using 2 (μg each of the wild type or mutant replicon DNA containing a partial tandem copy of pMPA1 or pMPA2 (Chatterji, A., *et al., J. Virol 73*:5541-5549 (1999)). The mutant replicon contained sequence alterations in the iteron sequences of the Rep protein in the viral origin. Total DNA from the protoplasts was extracted 48 hours after transfection (Dellaporta, S.L., *et al., Plant Mol. Biol. Reporter 1*:19-21 (1983); Mettler, I.J. *et al., Plant Mol. Biol. Rep. 5*:346-349 (1987)) and analyzed for viral DNA accumulation by Southern blotting (Chatterji, A., *et al., J. Virol 73*:5541-5549 (1999)).

***Results***

A. Expression and purification of the Rep protein from mild and severe strains of ToLCV

Previous studies (Chatterji, A., *et al., J. Virol 73*:5541-5549 (1999)) showed that the severe (pMPA1) and the mild (pMPA2) strains of ToLC-NdeV exhibit specificity in replication of their cognate DNA. This selectivity is determined by interaction between amino acid residues at the N-terminus of the AC1 and the Rep protein binding site in the intergenic region (IR). In this study, the nature of this interaction was directly examined *in vitro* by competitive DNA binding assays and *in vivo* by transient replication assays in tobacco protoplasts.

To isolate large amounts of Rep protein, the relevant sequences encoding the AC1 gene of the mild and the severe strains of ToLC-NdeV were expressed from a polyhedrin promoter in a baculovirus expression vector. The recombinant baculovirus was used to infect Sf9 cells to obtain a high titer virus stock and standardize optimal expression of the target protein. Cells were harvested at different time points after inoculation and protein was extracted after three cycles of freeze-thaw. The Rep protein from both the severe as well as the mild strain of ToLC-NdeV was detected in the crude insect cell lysates by immuno-precipitation using the polyclonal antiserum to AC1 gene (Fig. 2A, lanes 1, 2).

The high titer virus stock was used to infect Tn-5 cells for large-scale purification of the target protein. The soluble protein extracts were loaded on a Ni-NTA column and the eluted fractions were analyzed by SDS PAGE. The purified Rep protein had an estimated MW of 41 KD in coomassie stained polyacrylamide gels (Fig. 2B, lanes 2, 5) and its identity was further confirmed by immunoblotting using AC1 polyclonal antibody (Fig. 2C, lanes 1 to 7).

B. Rep proteins of the two strains bind a specific DNA sequence in the origin of replication

Based upon analogy with published reports describing binding sites of the Rep protein (Arguello-Astorga, G.R., *et al., Virology 203*:90-100 (1994); Fontes,

E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994a)), potential repeat motifs close to the TATA box in the common regions of mild and severe strains of ToLCV were identified. The ability of the Rep protein from the two strains to interact with their respective repeat motifs was determined in EMSAs. The potential repeat sequence in the severe strain was identified as GGTGTCTGGAGTC (nts 2640-2653) (U15015) (SEQ ID NO:121) while in the mild strain the repeat motif identified was GGCGTCTGGCGTC (nts 2640-2653) (SEQ ID NO:122 ) (U15017) (Chatterji, A., *et al., J. Virol 73*:5541-5549 (1999)). The 13-bp sequence (nts 2640-2653) containing the repeat motifs was used as the probe in EMSAs with the purified Rep protein. As a control, the 52-bp fragment of the common region (nts 2614 to 2666) from the respective strains was used as a probe in similar assays.

A distinct DNA-protein complex was observed using both the common region (52-bp) fragment as well as the 13-bp repeat sequence (Fig. 3, lanes 1 and 2). This data suggests that the 13-bp sequence may represent the binding site for the Rep protein of the severe strain. Similarly, the Rep protein of the mild strain formed a complex with its respective 52-bp common region fragment as well as the 13-bp repeat motif sequence (Fig. 3, lanes 6 and 7) from the common region.

The specificity of binding by the Rep proteins was tested in competition assays with radiolabelled probes and an excess of unlabeled DNA. Incubation of the Rep proteins with 50 fold molar excess of DNA of the same sequence abolished the formation of the complex completely (Fig. 3, lanes 3 and 8). On the other hand, a 1000 fold molar excess of an unrelated 13-bp sequence derived from the pUC 18 vector DNA did not affect the formation of the complex (Fig. 3, lanes 4 and 9). These results indicate that the interaction of Rep protein with its 13-bp sequence is highly specific. Since no obvious difference in terms of the mobility of the shifted complex was observed as a result of binding of the Rep protein to the 13-bp oligonucleotide or to the larger common region fragment containing the viral origin, the data suggested that the repeat sequences may constitute the high affinity binding site for the Rep protein.

C. Rep proteins of two strains exhibit specificity in binding to their cognate iterons

To determine whether the Rep protein from the mild strain could bind to the iteron sequence of the severe strain and vice-versa, purified proteins from the two strains were incubated with labeled 13-bp sequence containing the heterologous repeat motifs in EMSAs. No DNA-protein complexes were observed when the severe strain Rep protein was incubated with the oligonucleotide containing the binding site sequence of the mild strain (Fig. 3, lane 5). Furthermore, the mild strain Rep protein did not bind to the 13-bp sequence comprising the repeat motifs of the severe strain (Fig. 3, lane 10) of ToLC-NdeV.

D. Specificity of binding is related to the sequence, spacing and number of iterons

Unlike the iteron sequences of the mild strain of ToLC-NdeV, the sequence of iterons that constitute the binding site of the severe strain are not identical. To better understand the basic DNA sequence requirements that contribute to or influence specificity of origin recognition, chimeric iteron sequences were made by exchanging individual motifs in the binding site of the two strains and determined the ability of Rep protein to bind them. In addition, the effect of each of these mutations on virus multiplication was determined in transient assays by assessing the ability of mutant DNA–A components to replicate in tobacco protoplasts.

*Sequence of the repeat motif:* Since the iterons constituting the binding site are repeated, they are referred to as 5' or 3' depending upon their position in the common region. Mutant oligonucleotides were synthesized to determine whether the 5' and the 3' iterons contributed equally to the binding efficiency of the Rep protein and if binding specificity between the strains could be altered by exchanging the appropriate iterons. These oligonucleotides were used as probes in EMSAs to test their ability to form a complex with the purified Rep proteins.

*5' severe, 3' mild:* (IT ½ GGTGTCTGGCGTC) (SEQ ID NO:160) The Rep protein of the severe strain formed a strong complex with this sequence (Fig.

4A, lane 1). However, the Rep protein of the mild strain did not bind very efficiently to this probe resulting in a very weak complex (Fig. 4A, lane 7). In protoplasts, the severe strain background with this mutant origin was able to support replication of the virus to wild type levels but the mild strain accumulated very low levels of viral DNA (Fig. 5, lanes 2 and 13).

*5′ unrelated, 3′ severe* (IT ¾ GGGGTCTGGAGTC) (SEQ ID NO:182) ) Neither of the Rep proteins from the mild or the severe strain generated a complex with this probe *in vitro* (Fig. 4A, lanes 2 and 8). Also, in tobacco protoplasts, the accumulation of viral DNA was low (Fig. 5, lane 3) in the case of severe strain mutant but the mild strain mutant did not replicate any viral DNA at all (Fig. 5, lane 14)

*5′ unrelated, 3′ mild:* (IT 5/6 GGGGTCTGGCGTC) (SEQ ID NO:169) Neither of the Rep proteins bind this sequence (Fig.4A, lanes 6 and 10). Very low levels of viral DNA accumulation was detected in tobacco protoplasts when inoculated with the severe strain DNA–A component but the mild strain failed to replicate any viral DNA (Fig.5, lanes 4, 15).

*5′ mild, 3′ unrelated:* (IT 7/8 GGCGTCTGGGGTC) (SEQ ID NO:171 ) Neither of the Rep proteins formed a complex with this probe (Fig. 4A, lanes 6 and 10). In tobacco protoplasts, both mutants failed to accumulate viral DNA (Fig.5, lanes 5 and 16).

The iterons comprising the binding site of the severe strain Rep protein are not identical repeats. To determine if either of the iteron sequences influence the binding efficiency, synthetic oligonucleotides were designed with GGTGTCTGGTGTC (IT 9/10) (SEQ ID NO:165) and GGAGTCTGGAGTC (IT 11/12) (SEQ ID NO:193) as perfect repeats and tested in EMSAs for their capacity to bind the Rep protein of the severe strain. In gel shifts, the Rep protein bound to GGTGTCTGGTGTC (SEQ ID NO:194) as visualized by a retarded band but the binding to GGAGTCTGGAGTC (SEQ ID NO:195) was weaker in comparison to the mutant, 9/10. (Fig. 4A, lanes 3 and 4). In protoplasts, both mutants replicated viral DNA indistinguishable from the wild type controls (Fig. 5 lanes 6 and 7).

*Spacing within the iterons:* The iterated motifs in both the severe (GGTGTCTGGAGTC) (SEQ ID NO: 196) and the mild (GGCGTCTGGCGTC) (SEQ ID NO: 197) strain of ToLCV-Nde are separated by a single nucleotide. To find out if the spacing between the iterons was significant for origin recognition by the Rep protein, the distance between the iterons was either increased to six nucleotides (IT 13/14) or reduced to none by deleting the single T nucleotide (IT 15/16) between the iterons. DNA-protein complex were not observed when the spacing between the two repeats was either increased to six bases or decreased to zero (Fig. 3B lanes 1 and 2). Similar results were obtained with the Rep protein from the mild strain (Fig. 4B lanes 3 and 4). In protoplasts, neither of the two mutants accumulated viral DNA (Fig. 5, lanes 8, 9, 17, and 18).

*Number of iterons:* The Rep protein of the severe strain was unable to bind if only a single iteron (IT 17/18 GGTGTC) (SEQ ID NO:198)was used as a probe (Fig. 4B,lane 5). When the number of repeat motifs was doubled(IT19/20 GGTGTCTGGAGTCTGGTGTCTGGAGTC) (SEQ ID NO:199), multiple bands with retarded mobility were observed (Fig. 3B, lane 7). The Rep protein of the mild strain did not bind as efficiently to a monomer, GGCGTC (IT17/18). Doubling the number of repeat sequences (IT 19/20 GGCGTCTGGCGTCTGGCGTCTGGCGTC) (SEQ ID NO:200) did not improve efficiency of binding (Fig. 3B, lane 8). In transient assays, the virus mutant origins containing a single iteron did not replicate very well and only little accumulation of ss DNA (Fig. 5, lanes 10 and 19) was observed. But the mutants containing twice the number of repeat motifs in their origin were able to support virus replication and accumulated viral DNA similar to wild type levels (Fig. 5, lanes 11, and 20).

These results suggest that binding of the Rep protein to its cognate site is highly specific and the sequence of both the 5' and the 3' iteron is important for binding to occur. Secondly, while both the iterons are required for binding, they contribute differently to the efficiency of binding. The 5' iteron appears to be more important for binding, yet, substitution by an unrelated 3' iteron did not allow efficient complex formation. Third, the efficiency of binding of the Rep

protein to its cognate site can be correlated to the levels of virus replication and accumulation in protoplasts. Doubling the number of available binding sites generated multiple bands in gel shift assays but did not impact virus replication.

*Discussion*

This example defines DNA sequences in the viral origin of replication that are specifically recognized by the respective Rep protein of two strains of ToLC-NdeV and demonstrates that binding of the Rep protein to their cognate sequences is essential for viral replication. In addition, the binding of the Rep proteins to their cognate iterons is found to be highly specific between the strains and is dependent upon several criteria including the sequence, spacing and the number of iterons. Further, evidence is provided that any mutation in the iteron that affects DNA binding *in vitro* impacts viral DNA accumulation *in vivo*.

Previous studies (Chatterji, A., *et al.*, *J. Virol 73*:5541-5549 (1999)) identified 13-mer repeat sequences in the common region of ToLC-NdeV involved in interaction with the Rep protein and are essential for virus replication. Competitive DNA binding assays using purified Rep proteins established that the Rep protein of ToLC-NdeV (severe strain) specifically binds to the iterated sequence, 5' GGTGTCTGGAGTC (U15015) (SEQ ID NO:121) located on the 5' end of the TATA box between positions 2640 and 2653 and is conserved between the DNA-A and DNA-B components. Similar iterated sequences have been found in the common region of several geminiviruses (Arguello-Astorga, G.R., *et al.*, *Virology 203*:90-100 (1994)) and some have been verified in biochemical assays to act as Rep protein specific binding sites (Fontes, E.P.B., *et al.*, *Plant Cell 4*:597-608 (1992); Fontes, E.P.B., *et al.*, *J. Biol. Chem. 269*:8459-8465 (1994a); Behjatania, S.A., *et al.*, *Nucleic Acids Res. 26*:925-931(1998)). The specificity of binding was confirmed by competition assays. A 50 fold molar excess of the homologous probe abolished the formation of a DNA-protein complex, yet, a 1000 fold molar excess of heterologous DNA (pUC 18) did not affect the binding of the Rep protein indicating specificity of the

complex formed. The binding efficiency of the Rep protein to the 13 bp oligonucleotide or to the 52 bp common region fragment was indistinguishable. This data suggested that the complex observed as a result of Rep protein binding to the 13 bp sequence is authentic and the two repeat sequences represent the binding site of the Rep protein in both the mild and severe strains of ToLC-NdeV.

The sequence of the repeat motifs which constitute the binding site differ between the mild and the severe strain by two base pairs, GGTGTCTGGAGTC (U15015) (SEQ ID NO:121) for the severe strain versus GGCGTCTGGCGTC (U15017) (SEQ ID NO:122) for the mild strain. In EMSAs the Rep protein of the severe strain did not form a complex with either the 13-mer DNA sequence of the mild strain or the 52 bp fragment from the common region of the mild or severe strain, suggesting a high degree of specificity between the Rep protein of the two strains for their cognate sites. This observation is not surprising since there are several examples among geminiviruses documenting high specificity in replication of cognate genomes. As well, pseudorecombinants can only be formed between related viruses or strains of a specific virus (Lazarowitz, S.G., *et al., Plant Cell 4*:799-809 (1992); Stanley, J., *et al., Proc. Natl. Acad. Sci. U.S.A. 87*:6291-6295 (1990); Frischmuth, T., *et al., Virology 196*:666-673 (1993); Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994b)) reported that the Rep protein of bean golden mosaic virus (BGMV) can specifically recognize its cognate site but is unable to bind to the corresponding binding site of the related virus like TGMV. Similarly, the inability of Worland, CFH and the Logan strains of beet curly top virus (BCTV) to transcomplement replication of each other might be related to specific recognition of the binding site by the corresponding Rep proteins (Choi, I.R., and Stenger D.C., *Virology 206*:904-912 (1995); Stenger, D.C., *Phytopathology 88*:1174-1178 (1998)). Together, these results highlight the importance of specific interaction between the Rep protein and its cognate site in the replication of geminiviruses. In addition, it is important to emphasize the fact that precise recognition motifs of Rep proteins may be different between the viruses and are not interchangeable thereby maintaining specificity between closely related viruses and strains of the same virus.

Direct evidence for the essential role of Rep protein binding to the repeat sequences in geminivirus replication was obtained by the mutation of these sequences in the viral origin and testing their ability to replicate in tobacco protoplasts. It was found that the ability of the Rep protein to bind to its cognate site was correlated to the ability of the virus DNA to replicate in transient assays. Mutants that did not form a complex in the EMSAs did not replicate viral DNA suggesting that recognition or binding is a prerequisite for replication and viral DNA accumulation to occur.

The sequence requirements which govern the recognition specificity of the Rep protein for its cognate site was determined by altering sequence, spacing, and number of iterons. Mutagenesis of the Rep protein binding site showed that both repeat motifs are essential for DNA binding to occur, yet, the efficiency of binding was related to the sequence of its cognate 5' iteron. These results were also reflected in the reduced levels of virus accumulation in protoplasts transfected with mutant viral DNA suggesting that replication is directly correlated to the Rep protein binding to its specific site in the origin.

Furthermore, recognition of the binding site by the Rep protein was also found to be sensitive to any changes made in the spacing between the iterons. Drastically reduced levels of virus accumulation were observed in protoplasts when the spacing between the two iterons was changed to either six nucleotides or reduced to none. Because the Rep protein may bind as a dimer (Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994a); Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994b)), it is possible that the proximity of the two repeat motifs is congenial for binding to occur and any alterations with respect to spacing between the two iterons does not allow efficient recognition.

Our results indicate that the two iterons, even when they are identical make different contributions to the efficiency of binding as observed in the case of mild strain of ToLC-NdeV and has also been shown for TGMV (Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994a)). However, our results suggest that the 5' iteron is more important than the 3' one as far as binding is concerned, in case of TGMV (Fontes, E.P.B., *et al., J. Biol. Chem.*

*269*:8459-8465 (1994a); Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994b)) and BCTV(Choi, I.R., and Stenger D.C., *Virology 226*:22-126 (1996)), the 3' iteron has been shown to be more important for efficient binding as compared to the 5' repeat. In gel shift assays, the Rep protein of the severe strain was able to bind strongly to the sequence GGTGTCTGGTGTC (SEQ ID NO:201)than when GGAGTCTGGAGTC (SEQ ID NO:202) sequence was used as a probe. Also, the probe GGTGTCTGGTGTC (SEQ ID NO:203) was able to compete out any complex with the wild type repeat sequence, GGTGTCTGGAGTC (SEQ ID NO: 121) abolishing the binding of the severe strain Rep protein with its wild type iteron sequence. This data suggests that 5' iteron sequences are necessary for efficient binding of the Rep protein to its binding site.

The effect of deletion of a single base pair between the two repeat sequences resulted in negligible virus replication levels indicating that the ToLC-NdeV origin is sensitive to changes in the spacing of the iterons. Doubling the number of iterons resulted in the appearance of multiple bands which may indicate protein–protein interaction facilitated as a result of multimerization of the Rep protein, without causing any effect on accumulation of viral DNA. These results imply that recognition of cognate iterons may represent an essential step in replication process, yet, other interactions between these elements and possibly, yet-to-be-identified proteins that recognize or bind them.

***Summary***

The DNA binding sites for the replication associated protein (Rep) of two strains of tomato leaf curl virus from New Delhi (TolC-NdeV) were identified using electrophoretic mobility shift assays (EMSAs). The Rep proteins of the two strains were found to exhibit strict sequence specificity in recognition of their cognate repeat motifs (iterons) in the origin despite the fact that they share 91% sequence identity between them. Using a series of synthetic oligonucleotides as probes in EMSAs, the interaction of Rep protein with its binding site was found to be dependent upon number, size and sequence of the two iterons. Mutations in the sequence of the repeat motifs or alterations in the arrangement of the motifs

resulted in loss of DNA binding and accumulation of viral DNA in protoplasts suggesting that binding of Rep protein to its cognate iterons is an essential step in virus replication. In addition, a difference in sequence of two base pairs in the binding site of two ToLC-NdeV strains was found to affect DNA binding by the corresponding Rep protein and replication of the virus DNA in protoplasts.

***Example 3***
***Truncated Replication-Associated Proteins (Rep)***

Previous studies indicated a possible interaction between amino acid residues 9-10, specifically, amino acid 10 for ToLCV-Nde, at the N-terminus of the Rep protein and its binding site in the origin of replication that in turn determines specificity of replication between two strains. Mutagenesis and exchange of the 10th amino acid between the mild and the severe strain with concomitant change in the binding site sequence of ToLCV-Nde altered the replication ability of the two strains suggesting that these two components may influence the levels of viral replication and accumulation.

To better understand the significance of earlier observations which suggested an interaction between the N-terminal amino acid residues (9 and 10) of the Rep protein and the iteron sequences in the origin of replication, comparison was made of the ori and the N-terminal Rep protein sequences from a few geminiviruses, like Tomato golden mosaic virus (TGMV; iteron sequence, GGTAG), Bean golden mosaic virus (BGMV; iteron sequence, TGGAG), Tomato leaf curl virus (ToLCV-Nde: iteron sequence, GGTGT) and Potato yellow mosaic virus (PYMV; iteron sequence, GGGGG). Differences at amino acid positions 9 and 10 on the N-terminal of their Rep proteins were investigated. Differences were found in the viruses at position 10 as shown in Table 2.

The Rep sequence variation in the four cases stated in table 2 at amino acid 9 did not appear very significant (isoleucine and valine are both neutral amino acids), but at position 10 differences were observed with respect to different iterons. Since the arrangement, sequence and orientation of iterons

**Table 2. Comparison of iterons and the Rep N-terminal amino acids at position 9 and 10 in four different geminiviruses**

| Virus | $Rep^{9,10}$ | Iteron | Rep N-termianl Sequ |
|---|---|---|---|
| TGMV | I, N | GGTAG | FRIN |
| BGMV | V, Q | TGGAG | FRVQ |
| ToLCV-Nde | V, N | GGTGT | FRVN |
| PYMV | I, K | GGGGG | FSIK |
| PaLCV | I, N | GGGGA | FCIN |
| ITmLCV | I, N | GGTGG | FNIN |

within the members of the family geminiviridae reported so far are limited to only eight different types (Arguello-Astroga et al., 1994: listed in Figs. 1A-1C) one would expect that the variation at the N-terminus of Rep might be limited too. Results as outlined in the examples have shown that at least in case of ToLCV, these iterons bind the Rep protein specifically acting as Rep protein binding sites.

It has also been shown that binding is sequence specific so that the Rep protein from two highly homologous strains of ToLCV that differ in their iteron sequence binds only to cognate iterons. Results have demonstrated that viral replication is compromised if iteron sequences are exchanged between the strains without changing the Rep sequence. Taken together, these results strongly suggested that interaction between the Rep protein and its iteron sequence is the first crucial step in recognition and initiation of virus replication. Therefore, this can be exploited as a means to block or completely shut down viral replication in plants.

The ability of truncated and full length Rep proteins to compete for the repeat sequence (iterons) constituting the binding site of the protein was determined in vivo by transfecting tobacco protoplasts with 4 μg each of viral DNA-A and the truncated or full length rep protein on an expression vector.

Protoplasts were harvested 48h after inoculation and assayed for virus replication by analyzing total DNA on Southern blots.

Results suggest that in vivo, the truncated Rep proteins tend to compete with the full length Rep protein for the same binding site sequence, as evident by a reduction in the levels of viral DNA accumulation in case of both the mild and the severe strains. Co-inoculation of viral DNA-A with any of the three truncated Rep proteins appear to reduce virus accumulation but, in both mild and the severe strain Rep proteins, the T-Rep3 causes maximum reduction in the level of virus accumulation. Results of the transfections are provided in Tables 3 and 4.

**Table 3. Competition between the full length (A1 or A2) and truncated Rep proteins (T-Rep1, T-Rep2, T-Rep3) for the same iteron sequences**

| Virus | % replication |
|---|---|
| 1. A1 alone | 100 |
| 2. A2 alone | 100 |
| 3. A1+T-Rep1 | 90 |
| 4. A1+T-Rep2 | 45 |
| 5. A1+T-Rep3 | 48 |
| 6. A2+T-Rep1 | 40 |
| 7. A2+T-Rep2 | 50 |
| 8. A2+T-Rep3 | 5 |

Since this strategy is based upon a very fundamental step in virus infection and establishment in the host and because there appear to be limited types of iterons present within the Begomoviruses, it offers a broad application in control of geminiviruses by combining different types of truncated Rep proteins

recognizing different types of iterons. Based on this experiment, it is possible to compete with the wild type Rep protein by expressing a truncated Rep that will

**Table 4. Co-inoculation experiments to determine if the severe strain (A1) is dominant over the mild strain (A2) with respect to recognition of the cognate iteron**

| | Virus | % replication |
|---|---|---|
| 1. | A1+T-Rep1 of A2 | 50 |
| 2. | A1+T-Rep2 of A2 | 50 |
| 3. | A2+T-Rep1 of A1 | 5 |
| 4. | A2+T-Rep2 of A1 | 8 |
| 5. | A1+FLRep of A1 | 100- 125 |
| 6. | A1+FL Rep of A2 | 80 |
| 7. | A2+FL Rep of A2 | 100-120 |
| 8. | A2+FL Rep of A1 | 80 |

interfere with normal virus replication and in one case could reduce it to as low as 5% of the wild type levels.

## *Example 4*

### *Expression of N-Terminal Sequences of the Replication-Associated Protein (Rep)*

The binding sites of ToLCV-Nde Rep are not known. Potential binding site sequences were identified in the common region of ToLCV-Nde genome by site directed mutagenesis (Chatterji, A., *et al., J.Virol 73*:5481-5489 (1999)) and further confirmed by gel shift assays using purified Rep protein (Chatterji *et al.*, in preparation). The ToLCV-Nde Rep protein binds to the iterated motifs,

GGTGTCTGGAGTC (nts 2640-2653) (U15015) (SEQ ID NO:121) in the origin of replication. In this study the minimal binding domain on the AC1 gene of ToLCV was mapped by determining the ability of truncated Rep protein to bind origin DNA sequences using electrophoretic mobility shift assays (EMSAs). The effect of transient expression of truncated and full length AC1 sequences on viral DNA replication in *Nicotiana tabacum* BY2 protoplasts and *N. benthamiana* plants was also tested. Next, the ability of the truncated Rep protein of ToLCV-Nde to bind the iteron sequences of other geminiviruses in EMSAs was tested and therefore, the effect of its expression on replication of other geminiviruses was determined. The Rep protein of ToLCV-Nde specifically binds to its origin recognition sequence, GGTGTCTGGAGTC (SEQ ID NO:121) in EMSAs and the expression of its N-terminal sequences inhibits the replication of viral DNA. Further, transient expression of the ToLCV-Nde truncated Rep protein encoding the minimal binding domain could inhibit the replication of other geminiviruses that have similar iteron sequences.

**Materials and Methods**

*Plasmid Constructs*

A) Mapping the minimal binding domain of ToLCV AC1 gene.

Coding sequences corresponding to AC1 were PCR amplified and cloned in bacterial expression vector pET 28a (Novagen) and overexpressed in *E coli* cells. The recombinant proteins were named according to their C- and N-terminal amino acids. The C-terminal truncations were made by inserting an in-frame stop codon at positions 2436, pAC1-$1_{(1-52)}$; 2250, pAC1-$2_{(1-114)}$; and 2110, pAC1-$3_{(1-160)}$. The truncated AC1 sequences were sub-cloned as *Bam* HI to *Hind* III fragments in pET 28A vector digested similarly to give pAC1-1 and pAC1-2 and pAC1-3 respectively. At the N-terminus only one truncation was made by deleting the first 21 amino acids and inserting a *Nco* I site to create an in-frame start codon. The truncated fragment was cloned as a *Nco* I to *Hind* III fragment in the vector pET 28a to produce pAC1-$4_{(22-360)}$.

B) Construction of plant expression cassettes

For expression of a truncated AC1 gene in plant cells the C-terminal truncations described above were sub-cloned as *Bam* HI fragments in a similarly digested plant expression vector, plau 2 (pILTAB 350) under a cassava vein mosaic virus (CsVMV) promoter to produce the gene expression cassettes, pILTAB 401, pILTAB 402 and pILTAB 403 respectively. Constructions of infectious clones of pMPA1 and pMPB1 have been previously described (Padidam, M., *et al., J. Gen. Virol. 76*:25-35 (1995)). Full-length infectious dimers of African cassava mosaic virus (ACMV-kenya), pCLV 1.3A and pCLV 2B (Stanley, J., *Nature 305*:643-645 (1983)) and infectious monomers of pepper hausteco virus, PHV (Bonilla-Ramirez, G.M., *et al., J.Gen Virol. 78*:947-951 (1997)) were obtained. The potato yellow mosaic virus clones have been described (Umaharan *et al*, 1999).

C) Electrophoretic mobility shift assays (EMSA)

EMSA were performed similar to the method in Example 2

D) Transient Replication assays in protoplasts and plants

Transient replication assays in protoplasts were performed similar to Example 2

Transient replication assays in plants were performed as follows. Two week old seedlings of *N. benthamiana* were grown in magenta boxes and inoculated with partial tandem dimers of viral DNA using a Bio Rad helium driven particle gun (Padidam, M., *et al., J. Gen. Virol. 76*:25-35 (1995)). Ten plants were inoculated for each mutant using 0.5µg each of DNA-A and DNA-B genomic components per plant. Plants were observed for symptom development and the newly emerging leaves were harvested for Southern blot analysis after 3-4 weeks post inoculation.

E) Southern blot analysis

DNA extractions from systemically infected leaf samples were done as described in Dellaporta, S.L., *et al., Plant. Mol. Biol. 1*:19-21 (1983) and from protoplasts by following the procedure of Mettler (*Plant. Mol. biol. Rep. 5*:346-349 (1987). Total DNA (4$\mu$g) was electrophoresed on 1% agarose gels without ethidium bromide and transferred on nylon membranes. Viral DNA was detected using a component specific radiolabelled probe (a 900bp, AflII -Pst1 fragment containing the ORFs AC1, AC2 and AC3), or a probe specific for the B component (878 bp PCR amplified BC1 ORF). The amount of viral DNA was quantified as previously described (Chatterji, A., *et al., J.Virol 73*:5481-5489 (1999)) by exposing the Southern blots to phosphorscreens and counting on a phosphorimager (Molecular Dynamics).

F) DNA and polypeptide sequences or accession numbers used in the example.

Gene bank accession numbers of the DNA used in this study are as follows: pMPA1-U15015, pMPB1-U15017, pMPA2-U15016, ACMV-K- j02057, j02058, PHV-mex-X70418, 70419 and PYMV/TT-AF039031.

The peptide sequence of the full length Rep protein of ToLCV-Nde is:

MASPRRFRVNAKNYFLTYPKCSLTKEEALSQLQTLETPTKKKFIKICRE
LHEDGSPHIHVLIQFEGKFQCKNNRFFDLVSPSRSAHFHPNIQGAKSAS
DVKNYIAKDGDVLEWGVFQIDGRSARGGQQTANDAYAQAINTGNKD
DALRVLKELAPKDYVLQFHNLNTNLDRIFQPPSEVYVSPFSISSFDRVPA
DLVDWVSSNVVCAAARPFRPISIVIEGDSRTGKTMWARCLGPHNYLCG
HLDLSPKVYSNDAWYNVIDDVDPHYLKHFKEFMGAQRDCQSNTKYG
KPVMIKGGIPTIFLCNKGPNSSYKEYLDEEKNAALKQWAIKNAVFITLE
EPLYSGRENIALPEEEEEHSQEAS (SEQ ID NO: 183) (U15015)

**TRep1/pAC1-1** has the first 52 amino acids from the N-terminus cloned in plau 6 and the sequence is as follows:

MASPRRFRVNAKNYFLTYPKCSLTKEEALSQLQTLETPTKKKFIKICRE
LHE (SEQ ID NO:184) (U15015)

**TRep2/pAC1-2** has the first **114** amino acids from the N-terminus cloned in plau 6 and the sequence is as follows:

MASPRRFRVNAKNYFLTYPKCSLTKEEALSQLQTLETPTKKKFIKICRE
LHEDGSPHIHVLIQFEGKFQCKNNRFFDLVSPSRSAHFHPNIQGAKSAS
DVKNYIAKDGDVLEWG (SEQ ID NO:185) (U15015)

**TRep3/pAC1-3** has the first **160** amino acids from the N-terminus cloned in plau6 and the sequence is as follows:

MASPRRFRVNAKNYFLTYPKCSLTKEEALSQLQTLETPTKKKFIKICRE
LHEDGSPHIHVLIQFEGKFQCKNNRFFDLVSPSRSAHFHPNIQGAKSAS
DVKNYIAKDGDVLEWGVFQIDGRSARGGQQTANDAYAQAINTGNKD
DALRVLKELAPKDYVL (SEQ ID NO: 186 (U15015)

**Results**

A) Determination of the minimal binding domain of the Rep protein of ToLCv-Nde

The AC1 binds specifically to a directly repeated motif DNA sequence in the common region of the ToLCV-Nde genome. Purified Rep proteins were truncated at amino acids160, 114 and 52 to map the C-terminal boundary of the AC1 DNA binding domain *in vitro*. As a control, a full length Rep protein (encoding amino acids 1-360 of the AC1 gene) was used in all assays. The truncated and full length Rep proteins were over-expressed in *E. coli* under a T7 promoter and purified on a nickel affinity column. The affinity-purified proteins were highly enriched as determined by coomassie stained SDS PAGE gels and were detected in immuno-blots using the anti- histidine antibody.

His tagged AC1 proteins were tested for their ability to bind a radiolabeled 13 mer (nts 2640-2653) that includes the AC1 binding site sequence, 5'GGTGTCTGGAGTC3' (U15015) (SEQ ID NO:121). Shifted complexes were observed with full length (1-360) and C-terminal truncated proteins, pAC1-$3_{1\text{-}160}$. No binding was observed for pAC1-$1_{(1\text{-}52)}$ or pAC1-$2_{(1\text{-}114)}$. Together these results located the C-terminal boundary of the AC1 DNA binding domain to lie between amino acids 115 to 160. The N-terminal boundary of the AC1 binding domain

was determined *in vitro* by comparing the activity of the full length $AC1_{(1-360)}$ and N-terminal truncated $AC1_{(22-360)}$ to bind the 13 bp sequence, 5'GGTGTCTGGAGTC3' (U15015) (SEQ ID NO:121) in gel shift assays. The shifted complex was observed only with the full-length Rep protein and no bound DNA was detected for $AC1_{(22-360)}$. These results demonstrated that the sequences within the first 21 amino acids of AC1 are essential for protein DNA interactions. Put together, these results limited the DNA binding domain of ToLCV-Nde Rep protein to lie between amino acids 1-160.

B. ToLCV-Nde replication is inhibited by transiently expressed AC1

The effect of AC1 on viral DNA replication was investigated by co-inoculation of *N. tabacum* BY2 protoplasts with DNA-A and various expression cassettes expressing truncated AC1 gene sequences from a CsVMV promoter. ToLCV-Nde DNA-A alone replicated in BY-2 cells and accumulated

**Table 5: Virus replication in BY2 protoplasts and *Nicotiana benthamiana* plants co-inoculated with truncated Rep protein constructs and viral DNA (A1).**

| Virus construct | Symptom expression# | Replication* Protoplasts | Plants |
|---|---|---|---|
| A1+B | Severe, 10/10 | 100 | 100 |
| A2+B | Mild, 10/10 | 55 | 48-50 |
| A1+B+T-Rep1 | 9/10 mild, 1/10 severe | 100 | 90 |
| A1+B+T-Rep2 | Mild, 10/10 | 90 | 45 |
| A1+B+T-Rep3 | Mild, 10/10 | 30-40 | 45 |
| A1+B+T-Rep1/A2 | Severe, 10/10 | 100 | 94 |
| A1+B+T-Rep2/A2 | Severe, 10/10 | 100 | 89 |
| A1+B+T-Rep3/A2 | mild-intermediate, 10/10 | 60 | 45-50 |
| A2+B+T-Rep3/A2 | No symptoms, 10/10 | 50 | 56 |

#: A total of ten plants were inoculated. Shows the number of plants infected / number of plants inoculated. Plants were scored for symptom expression three weeks post inoculation.
*: The numbers refer to the amount (in percentage) of viral DNA replication in protoplasts electroporated with similar constructs. The viral DNA was quantified using a phosphorimager (Molecular Dynamics).

high levels of single stranded (ss) and supercoiled (sc) DNA (Fig 6A, lanes 1-4) but a significant decrease in the level of viral DNA replication (50-60% drop) occurred in the presence of truncated AC1 expressed from pAC1-3 (Fig 6B, lanes 1-12, Table 5).

Reduction in replication levels was estimated by quantifying the amount of radioactivity using a phosphorimager (Storm 860, Molecular Dynamics). The levels of reduction in viral replication was not as dramatic in the presence of expressed plasmids pAC1-1 and pAC1-2. pAC1-1 encodes 52 amino acids of the N-terminal of AC1 followed by an in-frame stop codon immediately after the 52 amino acids. pAC1-2 has the capacity to encode 114 N-terminal amino acids of the AC1 gene. Earlier experiments using the gel shift assays showed that the pAC1-1 and pAC1-2 do not contain an intact DNA binding domain as compared to pAC1-3, therefore, these results implied that presence of an intact DNA binding domain is essential to observe the inhibitory effect of truncated AC1 on viral replication.

Analogous truncations made in the AC1 gene of the mild strain of ToLCV-Nde exhibited similar inhibitory effect on viral DNA accumulation in BY2 protoplasts (Table 6).

**Table 6: Virus replication in BY2 protoplasts and *Nicotiana benthamiana* plants co-inoculated with truncated Rep protein constructs and viral DNA (A2).**

| Virus construct | Symptom expression# | Replication* Protoplasts | Plants |
|---|---|---|---|
| A1+B | Severe, 10/10 | 100 | 100 |
| A2+B | Mild, 10/10 | 55 | 52 |
| A2+B+T-Rep1 | Mild, 10/10 | 55 | 56 |
| A2+B+T-Rep2 | Mild, 10/10 | 45 | 42-45 |
| A2+B+T-Rep3 | Mild, 10/10 | 32-35 | 42 |
| A2+B+T-Rep1/A1 | mild, 10/10 | 50 | 60 |
| A2+B+T-Rep2/A1 | mild 10/10 | 52 | 58 |
| A2+B+T-Rep3/A1 | mild,10/10 | 50 | 58 |

C. Transformation of *N. benthamiana*

Two week old seedlings of *N. benthamiana* plants were co-bombarded with 2$\mu$g each of infectious dimers of DNA-A and DNA-B as well as truncated AC1 gene sequences expressed from CsVMV promoter. The plants were regularly screened for symptom development. Symptom severity was graded from asymptotic (score 0) to mild (very minor leaf curl, no puckering of leaves, no stunting of plants, score 1), intermediate (leaf curl symptoms but no adverse effect on growth, score 2) and severe (severe leaf curl, prominent blistering on the leaves and extreme stunting of plants, score 3). All (10/10) inoculated wild type plants developed severe symptoms five days after inoculation. In contrast, plants co-inoculated with pAC1-3 were less susceptible to ToLCV infection (Table 6) . Four out of ten plants were asymptomatic, 3/10 showed mild symptoms and only 2/10 plants expressed intermediate symptoms of leaf curl (Table 6). None of the 10 plants showed severe infection or stunted growth typically expressed by wild type inoculated plants. Most of the plants inoculated with pAC1-1 andAC1-2 developed severe symptoms seven days post inoculation (Table 6).

The level of viral DNA in ToLCV infected plants was analyzed by southern blot analysis of young leaves sampled 21days post inoculation using DNA-A (AC1) and DNA-B (BC1) specific probes (Fig 7). The viral DNA levels were variable ranging from undetectable to very low (an average of 15% of the WT levels) in asymptomatic plants but the accumulation of both genomic components increased with increasing severity of symptom expression. Plants co-bombarded with expression cassettes pAC1-1 and pAC1-2 developed intermediate to severe symptoms in majority of the cases.

D. Transiently expressed N-Rep of ToLCV reduces viral DNA accumulation in other geminiviruses

To further investigate the potential of truncated Rep protein to inhibit the replication of other geminiviruses, the effect of pAC1-3 expression on viral DNA accumulation of ACMV, PHV and PYMV/TT was studied. The rationale in choosing theses viruses was based on the sequence of their putative origin

recognition motifs present on their respective viral genome. If pAC1-3 binds to its homologous recognition sequence, GGTGTC (U15015) (SEQ ID N204O:205 ) in the origin, it might bind to similar or identical sequences even when they are present in the origin of other geminiviruses. Further, if binding to origin sequences is correlated to replication of the virus genome (as was observed in case of ToLCV-Nde), then it was predicted that as long as the truncated viral Rep pAC1-3 can bind to the origin function sequences of other geminiviruses it can potentially interfere in the replication of other geminiviruses as well.

The binding site sequences of ToLCV-Nde are not identical repeats, (GGTGTCTGGAGTC) (U15015) (SEQ ID NO:206). In case of ACMV, the putative iteron was identified as GGAGA (J02057) (SEQ ID NO:207); for PHV, the putative origin recognition motif may be GGTGA (SEQ ID NO: ) (X70418) and in case of PYMV, the potential binding sites may be GGTGT (SEQ ID NO: ) (AF039031) .

**Table 6: Regulation of virus DNA replication in BY2 protoplasts by the N-terminal sequences of AC1 gene of ToLCV-Nde**

| Virus | Iteron | N-Rep sequence | EMSA | pAC1-3 |
|---|---|---|---|---|
| ACMV | GGAGA | MRTPPRFRIQANKYFLTYPKC (SEQ ID NO: ) (J02057) | + | 33-37 |
| PHV | GGTGA | MPLPKRFRLNAKNYFLTYPQC (SEQ ID NO: ) (X70418) | +/- | 46-60 |
| PYMV | GGTGT | MP-PKRFRINANKYFLTYPKC (SEQ ID NO: ) (AF039031) | + | 47-50 |

Only T-Rep3 was tested in competition experiments. The numbers indicate virus replication levels as determined by southern blotting and phosphorimage analysis. Since any virus that had identical iteron sequences as the severe strain of ToLCV-Nde was not tested, it can only conclude that replication levels of the virus may go down in cases where competition is afforded by the truncated Rep. And that competition will result in cases where the origin sequences can be recognized by the truncated rep protein.

Co-inoculation of wild type DNA-A components of ACMV, PHV and PYMV with pAC1-3 of ToLCV-Nde Rep protein caused a decrease in the viral

replication levels (Table 6) even though the decrease in replication was not as significant as the inhibition observed in the case of pAC1-3 for its homologous ToLCV DNA.

*Discussion*

The minimal DNA binding domain of the AC1 gene of ToLCV to amino acids 1-160 was mapped and it was shown that the transient expression of these N-terminal sequences of the AC1 significantly inhibits ToLCV DNA accumulation in tobacco protoplasts and plants. It was found that the sequences comprising the first 1-52 or 1-114 amino acids of the Rep protein cannot effectively compete with the full length Rep protein to cause a reduction in virus accumulation even though a minor reduction in virus replication was observed. However, the pAC1-3 truncation afforded maximum competition to the full length Rep protein in terms of binding to the *ori* sequences as well as in reducing viral replication in protoplasts and plants. Comparison of competition experiments done between the mild and the severe strains of ToLCV-Nde in protoplasts and plants suggested that only homologous Rep sequences can compete and influence virus accumulation since the pAC1-3 from the severe strain did not affect the virus replication of the mild strain and vice versa.

Geminivirus Rep proteins are multifunctional and are involved in both replication and regulation of gene expression. The AC1 protein of TGMV has been known to bind with high affinity to repeat motifs located between the conserved TATA box and the initiation site of AC1 transcription (Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994a); Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994b)). It is presumed that this binding to the origin recognition sequence is responsible for the repression of AC1 transcription in TGMV (Sunter, G., *et al., Virology 195*:275-280 (1993); Eagle, P.A., *et al., Plant Cell 6*:1157-1170 (1994)) and ACMV (Hong, Y., and Stanley, J., *J.Gen.Virol. 76*:2415-2422 (1995)). On the basis of current understanding of Rep functions, there are several possible explanations for the down regulation of virus replication

levels by the AC1. Constitutive expression of the truncated viral rep protein could compete with the incoming Rep for binding to the viral origin of replication and influence viral accumulation levels thereby acting as a dominant negative mutant (Herskowitz, I., *Nature 329*:219-222 (1987)). Or it could adversely affect the integrity of the viral DNA by introducing nicks at cryptic motifs as described for wheat dwarf virus (Heyraud, F., *et al., EMBO J. 12*:4445-4452 (1993)). Another possibility is that the truncated Rep protein could repress transcription of the AC1 gene of the incoming virus by interacting with the upstream regulatory sequences in the origin. The NTP binding domain on the Rep protein of ToLCV was not mapped and it is possible that pAC1-3 lacks this domain and therefore inhibits viral DNA replication.

Based on the strength of the results obtained with homologous Rep and ori gin recognition sequences of ToLCV strains, the question is raised whether the same truncated Rep protein can interfere with the replication of related geminiviruses that have similar origin recognition sequences in their origin of replication. The results suggest that the transient expression of the ToLCV Rep protein encoding the DNA binding domain can reduce the replication of AVMV, PHV and PYMV to approximately 40-45% in tobacco protoplasts. Since the truncated Rep protein could also down regulate virus replication levels in other geminiviruses competition afforded by the pAC1-3 (competent for DNA binding) may be responsible for observed reduction in viral accumulation. Virus having identical iteron sequences as the severe strain of ToLCV-Nde was not tested, therefore it may be concluded that replication levels of the virus may go down in cases where competition is afforded by the truncated Rep.

Several approaches to control of geminiviruses have been developed. Transgenic *N. benthamiana* plants expressing defective interfering DNA (Stanley et al., 1990, Frishmuth, T., and Stanley, J., *Virology 183*:539-544 (1991)) of ACMV were less susceptible to ACMV infection but resistance was confined to closely related strains of ACMV owing to the need for AC1 mediated trans complementation of the DI DNA. Transgenic *N.tabacum* expressing antisense RNA targeted against TGMV AL1 (Day, A.G., *et al., Proc. Natl. Acad. Sci. USA*

*88*:6721-6725 (1991)) or tomato yellow leaf curl (Bendahamane and Gronenborn, 1997) showed that specificity of resistance depended on the level of homology between the antisense RNA and the target sequence. Finally, the potential of expressing a full length AC1 transgene in ACMV (Hong, Y., and Stanley, J., *Mol. Plant Microbe Int. 9*:219-225 (1996)) and the N-terminal sequences of TYLCV rep (Noris, E., *et al., Virology, 224*:130-138 (1996)) in virus resistance have also been observed. The current studies have not only extended our understanding of AC1 mediated resistance in terms of DNA-protein interactions but provided a means to exploit these interactions for achieving resistance in other geminiviruses that is not confined to related strains or viruses even when they belong to different geographical boundaries.

***Example 5***
***Identification of Replication Specificity Determinants***

Two strains of Tomato leaf curl virus from New Delhi (ToLCV-Nde) were used to further study specificity in replication of viral genomes. Rresults showed that the two strains specifically replicate their cognate DNAs and that specificity is determined in part by the amino acid at position 10 of the Rep protein and the corresponding binding site sequence in the *ori*. In addition, evidence is presented that the amino acid at position 10 may interact with the 3rd nucleotide of the 13-mer Rep protein binding site in the *ori*.

*Material and Methods*

A. Construction of Mutants

The cloning of DNA-A (pMPA1; accession no., U15015) and DNA-B (pMPB; Accession no., U15017) of the severe strain and DNA-A (pMPA2; accession no., U15016) of the mild strain have been described previously (Padidam, M., *et al., J. Gen. Virol. 76*:25-35 (1995)). The genome organization of the severe strain is shown in Figure 1A. DNA-A of the mild strain has the same

genome organization as DNA-A of the severe strain. The mutations analyzed in this study were targeted in the Rep gene and the common region of the viral genome. A brief description of each mutant used in this study and the method of construction is provided in Table 7. Full length Rep gene or its fragments were exchanged between the mild and the severe strains by utilizing the available restriction enzyme sites. In cases where convenient restriction sites were not present, oligonucleotide-directed mutagenesis (Horton, R.M., "*In vitro* recombination and mutagenesis of DNA," in *PCR cloning protocls., vol. 67*, White, B.A., ed., Humana Press, Inc., Totowa, NJ (1994), pp. 141-150) was used to substitute the fragments. For creating amino acid substitutions in the Rep protein, mutagenic oligonucleotides were designed to substitute codons. All mutants were confirmed by DNA sequencing. In case of substitutions made by oligonucleotide directed mutagenesis, a small restriction fragment containing the mutation was recloned into the unmutagenised A component at the respective site to avoid incorporation of second site mutations. Partial tandem dimers of the mutants were used to infect *Nicotiana tabacum* protoplasts and *N. benthamiana*.plants.

B. Protoplast and Plant Inoculations

Protoplasts derived from *N. tabacum* BY-2 suspension cultures were used for transfection with viral DNA (Watanbe, *FEBS Letters 219*:65-69 (1989)). One million protoplasts were inoculated by electroporation (250 V, 500 μF) with 2 mg each of DNA-A and DNA-B and 40 mg of sheared herring sperm DNA. Protoplasts were collected from cultures 48h post inoculation for DNA isolation and analysis.

**Table 7. Description of mutants used in this study**

| Construct | Method of construction |
|---|---|
| A2-RepA1 | DNA-A of mild strain containing the full length Rep gene from the severe strain. An *Nco* I site was introduced at the initiation codon of both the mild and the severe strain to facilitate the exchange of the AC1 gene. Full length Rep from the severe strain was digested with *Nco* I and *Bcl* I and cloned at respective site in the mild strain. |
| A2-cRepA1 | DNA-A of the mild strain containing the 3′ sequences coding for 256 amino acids from the C-terminal region derived from the severe strain. The C-terminal amino acids were cloned as *Cla* I to *Cla* I fragment from the severe strain. |
| A2-nRepA1 | DNA-A of the mild strain containing the 5′ sequences coding for amino acids 1-110 from the N-terminal region of Rep gene. Constructed by cloning the *Nco* I to *Xba* I fragment (coding for aa 1-110) from the severe strain into the mild at same sites. |
| A2-CRB1 | DNA-A of the mild strain containing the common region of severe strain DNA B. The common region of DNA B was amplified as *Xba* I to *Spe* I fragment. The primers designed to amplify common region from DNA-B had 18 nucleotides of the mild strain DNA-A at their ends in addition to DNA-B specific sequences. These sequences of the mild strain were later used as primers to extend and exchange the CR of A2. |

| Construct | Method of construction |
|---|---|
| A2-RepM1/CRA1 | A double mutant of the mild strain containing two amino acid substitutions, Val9 and Asp10 to Asn in Rep gene and the CR of A1. Amino acid substitutions were made by oligonucleotide-directed mutagenesis. |
| A2RepM1/CRB1 | A double mutant of the mild strain containing a mutated Rep gene (Val9 to Ile and Asp10 to Asn) and the CR of B1. |
| A2-RepM1/CRM3 | A double mutant of the mild strain containing Val9 to Ile and Asp10 to Asn in Rep gene as well as two point mutations in its CR. The point mutations in CR were made at positions 2642 (C to A) and 2649 (C to T) to make the binding site identical to A1. |
| A2-RepM2/CRM3 | A double mutant of the mild strain containing only Asp10 to Asn in Rep gene and the two point mutations in its CR at positions 2642 (C to A) and 2649 (C to T) respectively, rendering the binding site identical to A1. |
| A1-RepA2 | DNA-A of the severe strain containing the full length Rep from the mild strain. The Rep gene was isolated from the mild strain as *Nco* I to *Bcl* I fragment and cloned at same sites in A1. |
| A1-cRepA2 | DNA-A of the severe strain containing 3′ sequences coding for 256 amino acids from the C-terminal region of the Rep gene from the mild strain. The fragment was cloned between the two *Cla* I sites. |
| A1-nRepA2 | DNA-A of the severe strain containing 5′ sequences coding for amino acids 1 to 110 in the N-terminal region of the Rep gene from the mild strain. The fragment was clined as *Nco* I to *Xba* I between the strains. |

| Construct | Method of construction |
|---|---|
| A1-RepM3 | DNA-A of the severe strain containing a substitution at Asn10 to Asp in Rep gene. |
| A1-CRM1 | DNA-A of the severe strain carrying a single nucleotide deletion in the CR at position 2642. The wild type sequence of the putative binding site GGTGTCTGGAGTC is changed to GGGTCTGGAGTC due to this deletion. |
| A1-CRM2 | DNA-A of the severe strain containing a single nucleotide deletion in the CR at position 2649. The wild type sequence in the putative binding site, GGTGTCTGGAGTC is changed to GGTGTCTGGGTC due to this deletion. |
| A1-CRM4 | DNA-A of the severe strain containing a substitution of 3rd nucleotide in the putative binding site sequence GGTGTCTGGAGTC at 2642 GGCGTCGGAGTC. |
| A1-RepM4/CRM4 | DNA-A of the severe strain containing substitution of Asn10 to Asp in Rep gene and a single nucleotide change in the binding site sequence at 2642, making the repeat motif as GGCGTCGGAGTC. |

Two weeks old seedlings of *N. benthamiana* grown in magenta boxes were inoculated with partial tandem dimers of viral DNA using a Bio Rad helium driven particle gun (Padidam, M., *et al., J. Gen. Virol. 76*:25-35 (1995)). Ten plants were inoculated for each mutant using 0.5 mg each of DNA-A and DNA-B per plant. Plants were observed for symptom development and the newly emerging leaves were harvested for Southern blot analysis 21 days after inoculation.

B. Southern Blot Analysis

DNA extractions from systemically infected leaf samples were carried out as described in Dellaporta *et al.*, (Dellaporta, S.J., *et al.*, *Plant Mol. Biol. 1*:19-21 (1983)), and from protoplasts by following the procedure of Mettler (Mettler, I.J., *Plant Mol. Biol. Rep. 5*:346-349 (1987)). Total DNA (4 μg) was electrophoresed on 1% agarose gels without ethidium bromide and transferred to nylon membranes. Viral DNA was detected using DNA-A specific radiolabelled probe (a 900 bp, *Afl* II-*Pst* I fragment containing the Rep, REn and TrAP genes), or a probe specific for the DNA-B (878 bp PCR amplified movement protein gene). The amount of viral DNA was quantified as previously described (Padidam, M., *et al.*, *Virology 224*:390-404 (1996)) by exposing the Southern blots to phosphor screens and counting the radioactivity on a phosphorimager (Molecular Dynamics).

*Results*

A. The Mild Strain Does Not Replicate Efficiently

The two strains of ToLCV-Nde used in this study have been described previously (Padidam, M., *et al.*, *J. Gen. Virol. 76*:25-35 (1995)). The DNA-A (pMPA1) and DNA-B (pMPB) of the severe strain and DNA-A (pMPA2) of the mild strain have been cloned. All the three clones were obtained from the same DNA sample prepared from collection of several diseased plants in the same field. Inoculation of *N. benthamiana* plants with DNA-A and DNA-B of the severe strain produced severe symptoms with characteristic leaf curl but plants inoculated with DNA-A of the mild strain and DNA-B of the severe strain developed mild infection with minor leaf curl symptoms. For the sake of brevity, hereafter, the severe strain DNA-A and DNA-B will be referred to as A1 and B1 and the mild strain DNA-A will be designated as the A2. A1 and A2 DNAs have the same length (2739 nt.) and share 94% sequence identity. Their CRs are 81% identical

(Padidam, M., *et al., J. Gen. Virol. 76*:25-35 (1995)). The amino acid sequence identity between individual genes in A1 and A2 ranged from 91-99% with the greatest similarity in the coat protein gene. The nucleotide sequence identity between the CRs of A1 and B1 is 97% as compared to 79% between the CRs of A2 and B1. It was not known whether the mild phenotype in plants inoculated with A2 and B1 is due to inefficient replication of the virus or because of its inability to spread in the plant. To compare the replication levels of the two strains, BY-2 protoplasts were transfected with A1 or A2 DNAs and viral DNA replication was quantitated 48h after transfection.

A2 does not accumulate to the same level as the A1 in protoplasts. The replication efficiency of A2 DNA varied between 45-58% compared to the A1 DNA levels (Table 8). Next the ability of A2 to replicate B1 in tobacco protoplasts was tested. In transient assays, A2 replicated B1 to barely detectable levels (<1%) (Fig. 9B, lanes 1, 10; Table 8). In *N. benthamiana* plants inoculated with A2 and B1 DNA, very mild symptoms mostly limited to mild chlorosis and slight curling were observed three weeks post inoculation. Southern analysis of total DNA isolated from infected plants showed very low levels (5-10% compared to the A1) of DNA-B accumulation (Fig. 10A, lanes 1, 8 and 3B, lanes 1, 10; Table 8). In plants, replication of DNA B is required to cause a systemic infection (Padidam, M., *et al., J. Gen. Virol. 76*:25-35 (1995)) and these data suggested that low levels of DNA-B replication coupled with less efficient replication of A2 may account for mild symptoms observed in inoculated plants. The apparent inability of A2 to replicate B1 of the severe strain provided the first evidence that the two strains may have different replication requirements. Therefore, the incompatibility of A2 and B1 was focused on by making sequence comparisons between the Rep gene of A1 and A2 and the CR sequences between the A2 and B1.

B. N-terminal Domains of Rep Protein Are Not Interchangeable Between the Strains

The low levels of DNA-B accumulation in protoplasts and plants inoculated with A2 and B1 indicated that the mild strain Rep protein replicated B1 DNA inefficiently. To test if replication levels of DNA-B can be increased by replacing the mild strain Rep protein with its equivalent from A1, a full length Rep gene (from *Nco* I to *Bcl* I) was exchanged between the two strains (Fig. 8B, Table 7). The mild strain DNA-A containing a full length Rep gene of the severe strain (A2-RepA1) did not replicate efficiently in tobacco protoplasts (Fig. 9B, lanes 2, 11; Table 8). Similarly, the severe strain DNA-A containing the Rep gene from the mild strain (A1-RepA2) failed to accumulate to high levels in protoplasts (Fig. 9A, lanes, 2, 11; Table 8) suggesting that the Rep proteins are not interchangeable. Similar results were obtained in *N. benthamiana* plants inoculated with these mutants (Fig. 10A, lanes 2, 9 and 3B, lanes 2, 11; Table 8).

To determine whether the specificity of the Rep protein for the DNA templates was associated with C- or N-terminal regions of the Rep protein, we exchanged both the 5′ and the 3′ parts of the Rep gene between the strains. The 3′ region of the Rep gene coding for 256 amino acids (containing 18 of the 22 amino acid differences between the two Rep proteins) from the carboxyl terminal end of Rep (*Cla* I-*Cla* I fragment) was exchanged between the strains (A2-cRepA1 and A1-cRepA2; Fig. 8B; Table 7) and determined the ability of the Rep chimera to replicate in tobacco protoplasts. Unlike the exchange of full length Rep genes, the hybrid Rep proteins were functional in both strains but did not change the phenotype of the two strains. The severe strain mutant, A1-cRepA2 accumulated both DNA-A and DNA-B at levels similar to the wild type (A1) virus (Fig. 9A, lanes 3, 12; Table 8). In contrast, the mild strain mutant, A2-cRepA1 replicated DNA-A at moderate levels and very low levels of DNA-B was detected (Fig. 9B, lanes 3, 12; Table 8). Inoculation of *N. benthamiana* seedlings with these hybrid Rep gene constructs showed that the plants infected with the A2-cRepA1 produced mild symptoms while plants inoculated with A1-cRepA2

developed typical leaf curl symptoms within ten days (Table 8; Fig. 10A, lanes 3, 10 and 10B, lanes 3, 12).

It should be noted that by exchanging the full length Rep gene between the strains the overlapping TrAP gene was also transferred; however, the results described above eliminated the possibility that differences in the TrAP gene having an effect on replication of the chimeric viruses.

Sequences from the 5′ region of the Rep gene encoding amino acids 1-110 (*Nco* I-*Xba* I fragment) between the strains were exchanged and assayed for replication of viral DNA. The severe strain mutant, A1-nRepA2 contains the N-terminal sequences from the mild strain and led to accumulation of very low levels of viral DNA (Fig. 9A, lanes 4, 13; Table 8). Similarly, the mutant A2-cRepA1 resulted in negligible levels of virus replication (Fig. 9B, lanes 4, 13; Table 8). None of the *N. benthamiana* plants inoculated with either A1-nRepA2 or A2-nRepA1 mutants developed symptoms and accumulated very low levels of viral DNA (Table 8; Fig. 10A, lanes 4, 11 and 10B, lanes 4, 13) indicating the region spanning amino acids 1-110 in the Rep gene may contain residues that determine specificity of replication between the two strains and are not interchangeable.

C. Exhange of the intergenic region (IR) between the virus strains.

The IRs of pMPA1 and pMPA2 are only 80% identical while those of pMPA1 and pMPB are 97% identical. Further, pMPA2 and pMPB share only 77% sequence homology. It is not known if these differences in the IR were significant in determining the replication specificity observed earlier in the experiments and so, the IR of the pMPA2 (mild) was replaced with that of pMPB. This exchange however, did not improve the replication levels of DNA-B and the plants remained very mildly symptomatic. Infact, both ss and ds forms of DNA-A

**Table 8. Replication and Infectivity of ToLCV mutants in *N. tabacum* protoplasts and *N. benthamiana* plants**

| | Protoplast inocluations[a] | | | | Plant inoculations[b] | | | |
|---|---|---|---|---|---|---|---|---|
| | DNA-A | | DNA-B | | DNA-A | | DNA-B | |
| MUTANT[#] | ss | ds | ss | ds | ss | ds | ss | ds |
| A1[c] | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| A2 | 41 | 44 | <1 | <1* | 58 | 55 | 4 | 2 |
| A2-RepA1 | 4 | <1 | <1 | <1 | 5 | <1 | <1 | <1 |
| A1-RepA2 | <1 | <1 | <1 | <1 | 12 | 5 | <1 | <1 |
| A2-cRepA1 | 39 | 33 | <1 | <1 | 59 | 54 | <1 | <1 |
| A1-cRepA2 | 94 | 89 | 92 | 91 | 89 | 84 | 94 | 91 |
| A2-nRepA1 | 5 | <1 | <1 | <1 | 4 | <1 | <1 | <1 |
| A1-nRepA2 | <1 | <1 | <1 | <1 | 6 | <1 | 6 | 5 |
| A2-CRB1 | <1 | <1 | <1 | <1 | <1 | <1 | <1 | <1 |
| A2-RepM1/CRA1 | 91 | 92 | 84 | 82 | 94 | 92 | 98 | 97 |
| A2-RepM1/CRB1 | 92 | 84 | 90 | 87 | 92 | 89 | 106 | 104 |
| A2-RepM1/CRM3 | 102 | 96 | 94 | 97 | 98 | 95 | 104 | 98 |
| A2-RepM2/CRM3 | 106 | 98 | 89 | 84 | 92 | 79 | 85 | 71 |
| A1-RepM3 | 4 | 5 | <1 | <1 | 14 | 8 | <1 | <1 |
| A1-CRM1 | <1 | <1 | <1 | <1 | 3 | 2 | <1 | <1 |
| A1-CRM2 | 4 | 3 | 8 | 6 | <1 | <1 | <1 | <1 |
| A1-CRM4 | 12 | <1 | <1 | <1 | nd[d] | nd | nd | nd |
| A1-RepM4/CRM4 | 98 | 92 | 104 | 108 | nd | nd | nd | nd |

[#] RepM and CRM denote mutations in the Rep gene or in the CR respectively.

[a] The values shown are average (%) amounts of single stranded (ss) and double stranded (ds) viral DNA detected in sixteen independent protoplasts transfections per mutant. Protoplasts prepared from *N. tabacum* BY2 cells were transfected with 2 µg each of DNA-A and DNA-B and harvested 48th after electroporation. Viral DNA was quantitated using a phosphorimager. Standard error values between different transfections were in the range of ±2-5%.

[b] The values show average amount of viral DNA in ten inoculated *N. benthamiana* plants per mutant. The plants were inoculated with 0.5 µg each of DNA-A and DNA-B using a particle acceleration gun. Standard error values ranged from 2-5% between different plants.

[c] The amounts of viral DNA observed in protoplasts and plants inoculated with the severe strain were assigned a value of 100.

[d] Not determined.

[*] Too low for accurate quantification because of background error.

were sharply reduced to the limits of detection. These results established that the AC1 protein of the two strains displays strict specificity in recognising their respective replication origins and the interaction of the two may be important in driving strain specific replication.

D. Rep Proteins Display Specificity in Recognizing Replication Origins

Geminivirus replication requires a functional interaction of Rep protein with specific sequences in the *ori* (Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994); Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994)). Studies described above identified the N-terminal region of Rep protein as being important in determining replication of the two strains. The common region (CR) of the three DNA components, A1, A2 and B1 was compared to look for differences that may contribute to template specificity. The CRs of A1 and A2 are 81% identical, while the CRs of A2 and B1 share only 79% sequence homology; by comparison, A1 and B1 are 97% identical. To determine whether sequence differences in the CR of mild strain restricts its replication, the CR of A2 was replaced with that of B1 (mutant A2-CRB1; Table 1) and tested for its ability to replicate in BY-2 protoplasts and *N. benthamiana* plants. As shown in Table 8, exchange of the common region of A2 with B1 did not increase the replication levels of DNA-B

in protoplasts. Rather, the replication levels of both DNA-A and DNA-B were drastically reduced (Fig. 9B, lanes, 5, 14; Table 8). Similarly, this mutant did not replicate in *N. benthamiana* (Fig. 10B, lanes 5, 14; Table 8). These results demonstrated that the Rep protein of A2 does not recognize the common region sequences of B1 suggesting that the two strains specifically recognize unique sequences in their origin of replication.

The experiments provided information to delimit two essential features that influence replication specificity of two strains, the common region sequences and the N-terminal residues in Rep protein. Mutations in the N-terminal region (amino acids 1-110) of A1 and A2 Rep proteins were next introduced with concomitant changes in their viral origin of replication to analyze the precise determinants of a functional, replication competent interaction.

The Rep protein in geminiviruses shares sequence similarities with other initiator proteins that follow rolling circle mode of replication. Based on comparison of sequences of these proteins, Koonin and Ilyina (Koonin, E.V. and Ilyina, J.V., *J. Gen. Virol. 73*:2763-2766 (1992)) identified a domain amongst geminiviruses in the N-terminal region of Rep protein that may be involved in initiating rolling circle replication. In this region, at least three motifs have been identified: Motif III, [xxYxxK] is involved in DNA nicking and closing activities (Laufs, J.S., *et al., FEBS Lett. 377*:258-262 (1995); Hoogstraten, R.A., *et al., Mol. Plant. Microbe Interactions 9*:594-599), Motif II, HIHxUUQ (U= a bulky hydrophobic residue) which has structural features similar to the Mg2+binding sites of metallozymes (Koonin, E.V. and Ilyina, J.V., *J. Gen. Virol. 73*:2763-2766 (1992)) and Motif I, FLTYPqC (q = a basic or a polar amino acid) whose function has not been established yet. The three motifs described are present within the amino acids 1-110 region and are identical between the Rep proteins of A1 and A2.

To define the amino acids involved in recognition of the *ori*, the region between amino acids 1-110 was examined for sequence variation between Rep proteins of A1 and A2. Four amino acid differences were identified between the two proteins in this region. The A1 strain contains amino acids Ile, Asn, Lys, and

Glu at positions 9, 10, 40 and 52 respectively while the A2 strain has Val, Asp, Ala, and Asp at these positions. Two amino acids, Ile9 and Asn10 are immediately adjacent to Motif I. To determine if Ile9 and Asn10 had a role in replication, Val9 and Asp10 in Rep protein of A2 were changed to Ile9 and Asn10. Simultaneously, the common region of the A2 DNA was replaced with either A1 (A2-RepM1/CRA1) or B1 (A2-RepM1/CRB1) sequences (Fig. 10C). In protoplasts, both mutants, A2-RepM1/CRA1 (Fig. 9B, lanes 6, 15) and A2-RepM1/CRB1 (Fig. 9B, lanes 7, 16) accumulated viral DNA to similar levels as the severe A1 strain. *N. benthamiana* plants inoculated with both of these mutants developed severe infection ten days after inoculation and accumulated high levels of viral DNA as analyzed by Southern hybridization (Fig. 10B, lanes 6, 7, 15, 16; Table 8). These data suggested that Ile9 and Asn10 are involved in determining interaction of Rep protein with specific sequences in the CR.

E. Putative Rep Binding Sites Are Different Between the Two Strains

Rep protein in geminiviruses is known to bind with high affinity to its binding site in the *ori*. Based on comparison of many *ori* sequences in geminiviruses (Arguello-Astorga, G.R., *et al., Virology 203*:90-100 (1994); Behjatnia, S.A., *et al., Nucleic Acids Res. 26*:925-931 (1998); Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994)) putative Rep binding sites in the CRs of A1, A2 and B1 DNAs.were looked for. A 13-bp sequence was identified in *ori* close to the TATA box in the CRs of the three DNAs. In A1 and B1, the repeat sequence GGTGTCTGGAGTC (SEQ ID NO: 121) was identified while A2 DNA had the repeat sequence GGCGTCTGGCGTC (SEQ ID NO: 122).

To determine if these sequences represent potential binding sites for Rep proteins, we introduced mutations in the 13-bp sequence. Deletions were made at the 3rd or the 10th nucleotide of the 13-nucleotide sequence since they are different between the two strains. Both mutants, A1-CRM1 (deletion for the 3rd nucleotide) and A1-CRM2 (deletion for the 10th nucleotide) showed dramatically reduced levels of viral DNA replication in infected protoplasts (Fig. 9A, lanes 6,

7, 15, 16; Table 8). Inoculated *N. benthamiana* plants did not show any symptoms three weeks post inoculation and replication of both viruses was reduced to minimal levels (Fig. 10A, lanes 6, 7, 13, 14; Table 8). These results demonstrated that the 13-bp sequence in A1 is essential for virus replication and may represent the Rep protein binding site.

F. Asn10 May Determine Specific Interaction with the Viral Origin

Earlier experiments showed that amino acids Ile 9 and Asn10 in Rep protein of A1 may be involved in specific interaction with the CR. To determine if these amino acids are involved in recognition of the potential binding site in the *ori*, mutations were made in A2 Rep gene to change Val 9 to Ile and Asp10 to Asn. Simultaneously, the 3rd and the 10th nucleotides on the potential binding site in A2 (GGCGTCTGGCGTC) (SEQ ID NO: 122)were mutated to T and A respectively (GGTGTCGGAGTC) (SEQ ID NO: 121) to make the repeat sequence identical to that of A1 (mutant A2-RepM1/CRM3, Fig. 8C). As expected, the mutant A2-RepM1/CRM3 was functional in protoplasts and replicated DNA A and B to wild type levels (Fig. 9B, lanes 8, 17; Table 8). Also, plants inoculated with mutant A2-RepM1/CRM3 and B1 produced severe symptoms within two weeks after inoculation accumulating high levels of viral DNA ( Fig. 10B, lanes 8, 17; Table 8). These results implied that the 13-mer sequence identified in the *ori* of A2 is the putative binding site and plays a significant role in replication, and that Ile9 and Asn10 may be involved in specific interaction between Rep and *ori*.

Of the two amino acid changes made, Asp10 to Asn is expected to be more significant than Val9 to Ile9; we therefore made another mutant A2-RepM2/CRM3, where only the Asp10 to Asn of Rep protein in A2 is changed together with 3rd (C to T) and the 10th (C to A) nucleotides of the *ori* to make it identical to A1 (Fig. 10C). In an analogous experiment, the Asn10 to Asp was made in Rep protein of A1 (A1-RepM3) without any changes in its *ori* (Fig. 8C). Protoplasts transfected with A2-RepM2/CRM3 accumulated high levels of viral

DNA (Fig. 9B, Lanes 9, 18; Table 8) comparable to the wild type severe strain. The corresponding mutant, A1-RepM3 with Asn10 to Asp change in the Rep protein of A1, replicated to very low levels in protoplasts (Fig. 9A, lanes 5, 14; Table 8). *N. benthamiana* plants inoculated with the mutant A2-RepM2/CRm3 developed severe symptoms and accumulated increased levels of viral DNA (Fig. 10B, lanes 9, 18; Table 8), showing that Asn10 alone is sufficient to facilitate specific replication of the virus. In contrast, plants inoculated with A1-RepM3 which contains a substitution in the 10th residue of its Rep protein did not replicate viral DNA with high efficiency (Fig. 10A, lanes 5, 12; Table 8). Asn10 may interact with the 3rd nucleotide of the 5' iteron, GGTGTC

While the above experiment showed that Asn10 in Rep protein may be involved in specific recognition of the *ori*, it was essential to determine if both the 3'and the 5' repeats in the 13-mer binding site contributed to the specificity of recognition. To address this question, the A1 DNA was mutated at the 3rd nucleotide in this sequence substituting C for T (A1-CRM4) so that the altered 5' iteron is GGCGTCTGGAGGTC. The mutant A1-CRM4 did not replicate efficiently in protoplasts and very low levels of viral DNA accumulated (Fig. 9A, lanes 8, 17; Table 8), suggesting that Rep protein may be unable to recognize the modified binding site. This mutant was further modified by substituting Asn10 to Asp (A1-RepM4/CRM4, Fig. 10C). The mutant A1-RepM4/CRM4 accumulated wild type levels of DNA-A in protoplasts (Fig. 9A, lanes, 9, 18; Table 8) indicating that the Asp10 may indeed interact with the 3rd base of the iteron, GGCGTC sequence.

Additionally, in order to better define the AC1 domain involved in specific recognition of origin, the region between aa 1-110 was scanned for functional variation in the two Rep proteins(data not shown). A stretch of ten amino acids was identified at the beginning of the AC1 very close to the conserved motif 1(Koonin and Illiyana,1991; Arguello-Astroga et al,1995) and switched this domain in pMPA2 already containing the pMPB IR. This double mutant seemed functional in protoplasts and both forms of viral DNA components could be detected in same amounts as the severe pMPA1 strain. These data demonstrated

that the aa1-10 at the AC1 n-terminal region must be provided in trans for specific recognition of its binding site on the IR and consequent replication of the viral DNA..

***Discussion***

The specificity of interactions between Rep proteins and *ori* sequences in two related strains of Tomato leaf curl virus-New Delhi was investigated. The studies showed that the amino acid at position 10 in Rep protein coupled with a change in the binding site sequence may determine whether or not the viral DNA is replicated. Substitution of Asp10 to Asn in Rep protein of the mild strain accompanied by exchange of the 13-mer binding site (making it identical to the severe strain) altered its replication leading to increased accumulation of viral DNA. In addition, the mild strain thus modified could replicate heterologous strain DNA-B indicating that the interaction of Rep protein with its binding site may be essential for replication of viral DNA. Based on site directed mutations, it is proposed that Asn10 specifically recognizes the 3rd base pair of the 5′ iteron GGTGTC in the severe strain.

Even though the DNA-A component of the severe (A1) and mild (A2) strains share 94% sequence identity, A2 did not replicate efficiently in protoplasts or plants, nor did it support the replication of DNA- B1. These results support the hypothesis that the mild symptoms in plants inoculated with A2 and B1 are caused by low levels of replication of DNA-A combined with very low levels of DNA-B. Comparison of amino acid sequences in Rep proteins of the two strains revealed 22 amino acid differences, 18 of which are located in the C-terminal region of Rep protein. The results of experiments in which sequences of Rep proteins were exchanged indicated that the region encoding 256 amino acid residues from the C-terminal end did not affect viral replication. On the contrary, exchange of amino acids 1 to 110 of Rep was deleterious to virus accumulation suggesting that this region may contain sequences crucial for virus replication. However, N-terminal sequences in Rep protein alone may not account for the replication

specificity was shown by the mild strain mutant A2-CRB1 which contained the exchanged CR but failed to replicate viral DNA. These results imply that Rep proteins exhibit specificity in their interaction with their template and are not interchangeable. Similar results have been obtained with strains of TYLCV (Jupin, I., *et al., FEBS Lett. 262*:116-120 (1995)), TGMV (Gladfelter, H.J., *et al., Virology 239*:186-197 (1997); Orozco, B.M., *et al., J. Biol. Chem. 272*:9840-9846 (1997)) and beet curly top virus, BCTV (Choi, I-R., and Stenger, D.C., *Virology 206*:904-912 (1995); Choi, I-R., and Stenger, D.C., *Virology 226*:22-126 (1996); Watanabe, Y., *et al., FEBS Lett 219*:65-69 (1987)).

Of the four amino acid differences in the N-terminal region (amino acids 1-110) between the Rep genes of A1 and A2, we chose to mutate Ile9 and Asn10 because of their proximity to Motif I (FLTYPKC), a conserved element found in all the initiator poteins that replicate via a rolling circle mechanism (Koonin, E.V. and Ilyina, J.V., *J. Gen. Virol. 73*:2763-2766 (1992)). Replication assays in protoplasts and in plants obtained with the mild strain mutants A2RepM1/CRA1, and A2-RepM1/CRB1 provided evidence that these amino acids may be involved in specificity because of their interaction with the sequences in the CR.

The binding site of Rep protein of ToLCV-Nde has not been biochemically determined. Site directed mutagenesis was used to determine whether or not the 13 nucleotide sequence identified in the *ori* interacts in a functional way with the Rep protein. The severe strain mutants A1-CRM1 and A1-CRM2 which contain single nucleotide deletions in the 13-mer sequence, failed to accumulate viral DNA demonstrating that the CR sequence is essential for virus replication. Since these deletions also affected spacing of the putative binding site, these results indicate that both sequence and spacing may contribute to specificity. Mutational analysis of the Rep binding site in TGMV showed that both spacing and sequence of the binding site are important for replication (Orozco, B.M., *et al., Virology 242*:346-356 (1998)).

The mutant A2-RepM2/CRM3 which contains an Asp10 to Asn mutation in the Rep protein and corresponding changes in the potential binding site sequence GGCGTCTGGCGTC (SEQ ID NO:122) to GGTGTCTGGAGTC

(SEQ ID NO:121) (identical to the severe strain) restored the replication efficiency of A2 DNA. These results indicated that Asn10 may differentiate between A1 and A2 strains and determine the specificity in recognizing *ori* sequences. In addition, the fact that replication of B1 was restored by changing the putative binding site sequence of A2 coupled with mutation of the Rep protein support the conclusion that both components are key factors that determine which DNA template is to undergo replication. The role of the other two amino acids Lys40 and Glu54, that are different between the two Rep proteins was not studied and therefore do not conclude that Asn 10 is solely responsible for strain specificity.

Since the iteron sequences in the binding sites of A1 and A2 are different with respect to the 3rd and the 10th nucleotides, we examined the significance of these differences in context of Asn10 in Rep protein. These studies led to the conclusion that a GGCGTC iteron in the putative binding site may be correlated with the presence of Asp in Rep protein at position 10 of A2. Likewise, in related experiments we showed that Asn10 may recognize the 3rd base pair in the 5′ iteron GGTGTC. It is possible that Asn10 is a part of the DNA-binding domain of Rep protein which allows appropriate structural presentation of the Rep protein that potentiates recognition with the iteron sequences by Rep protein and facilitates replication. In TGMV, the deletion of the first 29 amino acids abolished DNA binding and DNA cleavage demonstrating that an intact N-terminus is required for both activities (Orozco, B.M., *et al., J. Biol. Chem. 272*:9840-9846 (1997)). The inability of the Cal/Logan strain to replicate the Worland or the CFH strain of BCTV (Watanabe, Y., *et al., FEBS Lett 219*:65-69 (1987)) was correlated to the differences in the 3rd base pair of the 5' iteron sequence supporting our observations that variation in this region of the sequence may be crucial in determining the replication ability between the strains. Similar reports of incongruity have been shown for different strains of TYLCV (Jupin, I., *et al., FEBS Lett. 262*:116-120 (1995)) whose Rep proteins share more than 76% anmino acid identity but are not interchangeable and between TGMV and BGMV (Fontes, E.P.B., *et al., Plant Cell 6*:405-416 (1994)), probably beacuse of

differences in their iteron sequences, supporting the importance of specific interaction between the Rep protein and its recognition sequence.

The iteron sequences of A1 and B1 are not identical. Point mutations introduced in the 5′ iteron of the severe strain (mutant A1-CRM4) to make it identical to its 3′ homolog (A1-CRM4) resulted in drastic reduction in virus replication indicating that the two iterons do not contribute equally to the recognition process. Yet, a deletion of the 10th nucleotide of the repeat motif (A1-CRM2) reduced the replication levels in A1 suggesting that both iterons are required for replication. Similarly, differential contributions of the two iterons have been reported for TGMV (Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994)) and BCTV (Choi, I-R., and Stenger, D.C., *Virology 226*:22-126 (1996)). Unlike the reports of Fontes *et al.*, (Fontes, E.P.B., *et al., J. Biol. Chem. 269*:8459-8465 (1994)) and Choi and Stenger, (Choi, I-R., and Stenger, D.C., *Virology 226*:22-126 (1996)), our studies indicated that the 5′ iteron contributes to replication more than the 3′ iteron, for example, in mutant A1-CRM4. It is possible that in the case of ToLCV, the Rep protein has a stronger affinity for the 5′ versus 3′ iteron, but detailed *in vitro* binding assays that examine the interaction of Rep protein with each of the iterons are required to confirm this suggestion.

The studies showed a correlation between the amino acid at the 10th position in Rep protein and the 3rd nucleotide of the 5' iteron in the binding site and specificity of replication between the strains. This suggestion was supported by several observations: A2 DNA mutated at Asp10 to Asn in the Rep protein with concomitant change in the 3rd nucleotide of the putative binding site (A2-RepM2/CRM3) resulted in increased levels of virus accumulation. Similarly, the mutant, A1-RepM4, containing a substitution of Asn10 to Asp in Rep protein without any change in its binding site accumulated very low levels of virus DNA. Further, substitution of Asn10 to Asp accompanied by a change in the 5′ iteron of the binding site (mutant A1-RepM4/CRM4), restored virus replication suggesting that the correlation between the amino acid at position 10 and the 3rd nucleotide of the iteron may indeed be related to specificity of replication.

The observed specificity of the Rep protein with sequences in the *ori* accounts for selective replication of A1 and A2 strains. While earlier work (Choi, I-R., and Stenger, D.C., *Virology 206*:904-912 (1995); Choi, I-R., and Stenger, D.C., *Virology 226*:22-126 (1996); Jupin, I., *et al., FEBS Lett. 262*:116-120 (1995); Lazarowitz, S.G., *et al., Plant Cell 4*:799-809 (1992)) has shown that specificity of replication may reside within the N-terminal sequences of the Rep, our work delimits the specificity determinants to amino acid Asn10 of Rep protein in case of the two strains of ToLCV-Nde . Since the Asn10 is very closely associated with the conserved Motif I sequence, FLTYPKC (Koonin, E.V. and Ilyina, J.V., *J. Gen. Virol. 73*:2763-2766 (1992)) in Rep protein, we suggest that it may function as a part of the Motif I to mediate specific replication of the cognate genomes.

All publications, patents and patent applications cited herein are fully incorporated by reference into the disclosure.

Having now fully described the invention by way of illustration and example for purposes of clarity and understanding, it will be apparent to those of ordinary skill in the art that certain changes and modifications may be made in the disclosed embodiments., and such modifications are intended to be within the scope of the present invention.

## *What Is Claimed Is:*

1. A method for producing resistance in a plant to a geminivirus comprising introducing a geminivirus replication associated protein (Rep)-iteron antagonist into a plant, plant cell or propagule, wherein said antagonist is selected from the group consisting of a nucleotide sequence of a geminivirus iteron capable of binding to a Rep protein and a defective Rep, wherein said defective Rep comprises a conserved geminivirus iteron binding site.

2. The method of claim 1 wherein said Rep protein comprises an amino acid sequence selected from the group consisting of MPPPKKFRVQAKNYFLTYP (SEQ ID NO:1), MRTPRFRIQAKNVFLTYP (SEQ ID NO:2), MRTPRFRVQAKNVFLTYP (SEQ ID NO:3), MRTPRFRVQAKNVFLTYP (SEQ ID NO:4), MPPPKKFRVQSRNYFLTYP (SEQ ID NO:5), MPPPQRFRVQSKNYFLTYP (SEQ ID NO:6), MPPPKRFRINAKNYFLTYP (SEQ ID NO:7), MPPPQRFRVQSKNYFLTYP (SEQ ID NO:8), MAPPKRFKVQAKNYFITYP (SEQ ID NO:10), MAPPKRFKVQAKNYFITYP (SEQ ID NO:11), MAPPKRFKVQAKNYFITYP (SEQ ID NO:12), MPSKPRRFRVQAKNIFLTYP (SEQ ID NO:13), MPPPKKFRVQSKNHFLTYP (SEQ ID NO:14), MPPPKKFRVQSKNHFLTYP (SEQ ID NO:15), MPPPKRFRVNSKNYFLTYP (SEQ ID NO:16), MPPPKKFRVQSKNHFLTYP (SEQ ID NO:17), MPPPKRFRVNSKNYFLTYP (SEQ ID NO:18), MPPPKKFRVQSKNYFLTYP (SEQ ID NO:19), MRTPRFRIQAKNVFLTYP (SEQ ID NO:20), MAPPKRFKIYAKNYFLTYP (SEQ ID NO:21), MPPPKRFKINAKNYFLTYP (SEQ ID NO:22), MPRNPKSFRLAARNIFLTHL (SEQ ID NO:23), MPRALKTNAKNYFLTYP (SEQ ID NO:24), MPRSGAFRVNAKNIFATYP (SEQ ID NO:25), MPRLGRFAINAKNYFLTYP (SEQ ID NO:26), MPLPKRFRLNAKNYFLTYP (SEQ ID NO:27), MPPKRFRINSKNYFLTYP (SEQ ID NO:28), MPRNPNSFRLTARNIFLTYP (SEQ ID NO:29), MAPPNKFRINAKNYFLTYP (SEQ ID NO:30), MTRPKSFRINAKNYFLTYP (SEQ ID NO:31),

MPRPGRFNINAKNYFLTYP (SEQ ID NO:32), MASPRRFRVNAKNYFLTYP (SEQ ID NO:33), MPRAGRFSINARNYFLTYP (SEQ ID NO:34), MHPLNKFRINAKNYFLTYP (SEQ ID NO:35), MPRLFKIYAKNYFLTYP (SEQ ID NO:37), MPRLFKIYAKNYFLTYP (SEQ ID NO:38), MPRLFKIYAKNYFLTYP (SEQ ID NO:39), MPPSKKFLINAKNYFLTYP (SEQ ID NO:40), MPPSKKFLINAKNYFLTYP (SEQ ID NO:41), NRSFRHRNANTFLTYS (SEQ ID NO:43), SNRQFSHRNANTFLTYP (SEQ ID NO:44), SNRQFSHRNANTFLTYP (SEQ ID NO:45), SNRQFSHRNANTFLTYP (SEQ ID NO:46), SNRQFSHRNANTFLTYP (SEQ ID NO:47), SNRQFSHRNANTFLTYP (SEQ ID NO:48), SNRQFSHRNANTFLTYP (SEQ ID NO:49), SNRQFSHRNANTFLTYP (SEQ ID NO:54), SNRQFSHRNANTFLTYP (SEQ ID NO:55), SNRQFSHRNANTFLTYP (SEQ ID NO:56), SNRQFSHRNANTFLTYP (SEQ ID NO:57), SNRQFSHRNANTFLTYP (SEQ ID NO:59), SNRQFSHRNANTFLTYP (SEQ ID NO:60), SNRQFSHRNANTFLTYP (SEQ ID NO: 61), SNRQFSHRNANTFLTYP (SEQ ID NO:62), HSVRSFRHRNANTFLTYS (SEQ ID NO:64), HSVRSFRHRNANTFLTYS (SEQ ID NO:65), PSRRFKHRNVNTFLTYS (SEQ ID NO:66), TKSFRLQTKYVFLTYP (SEQ ID NO:67), PRFRVYSKYLFLTYP (SEQ ID NO:68), PRFRVYSKYLFLTYP (SEQ ID NO:69), MPPTKRFRIQAKNIFLTYPQ (SEQ ID NO:70), MPRTPKRFRIQAKNIFLTYPQ,(SEQ ID NO:71) MPFYKKAKNFFLTYPQ (SEQ ID NO:72), MPRSPSFQIKAKNIFLTYPR (SEQ ID NO:73), MPRQPNSFRIQARNIFLTYPQ (SEQ ID NO:74), MPSNPKRFQIAAKNYFLTYPN (SEQ ID NO:75) MPPPKRFQINSKNYFLTYP (SEQ ID NO:76), MAPPKRFQINAKNYFLTYP (SEQ ID NO:77), MPRAGRFQINAKNYFITYP (SEQ ID NO:78), MPRAGRFQINAKNYFVTYP (SEQ ID NO:79), MPRAGRFQINAKNYFITYP (SEQ ID NO:80), MSPPKRFQINAKNYFLTYP (SEQ ID NO:81), MAPPKQFQIYAKNYFITYP (SEQ ID NO:82), MPPKRFLINSKNYFLTYP (SEQ ID NO:83), MPSHPKRFQINAKNYFLTYP (SEQ ID NO:84),

MAPPKRFQINCKNYFLTYP (SEQ ID NO:85), MAQPKRFQINAKHYFLSFP (SEQ ID NO:86), MAQPKRFQINAKHYFLTFP (SEQ ID NO:87), MPRAGRFQINAKNYFITYP (SEQ ID NO:88), MPRNNSFCINAKNIFLTFP (SEQ ID NO:89), MPRNNSFCINAKNIFLTFP (SEQ ID NO:90), MPRLNSFCVNAKNIFLTYP (SEQ ID NO:91), MAAPNRFKLNAKNYFLTYP (SEQ ID NO:92), MPRKGSFSVKAKNYFLTYP (SEQ ID NO:93), MPPPKRFSVNAKNFFLTYP (SEQ ID NO:94), MPRKGSFSIKAKNYFLTYP (SEQ ID NO:95), MPRKGSFSIKAKNYFLTYP (SEQ ID NO:96), MPRKGYFSVKAKNYFLTYP (SEQ ID NO:97) and MPRSGRFSIKAKNYFLTYP (SEQ ID NO:98).

3. The method of claim 1 wherein said protein forms a dimer with wild type geminivirus Rep protein.

4. The method of claim 1 wherein said protein comprises from two to thirty different conserved iteron binding sites.

5. The method of claim 1 wherein said Rep protein includes the amino-terminal Rep protein sequence according to a formula selected from the group consisting of -FRVQ- (SEQ. ID NO:126), -FRVN- (SEQ. ID NO:127), -FRIN- (SEQ. ID NO:128), -FRIQ- (SEQ. ID NO:129), -FRLQ- (SEQ. ID NO:130), -FKVQ- (SEQ. ID NO:131), -FKIY-(SEQ. ID NO:132), -FKIN- (SEQ. ID NO:133), -FRLA- (SEQ. ID NO:134), FRLN- (SEQ. ID NO:135), -LKTN- (SEQ. ID NO:136), -FAIN-(SEQ. ID NO:137), -FRLT-(SEQ. ID NO:138), -FNIN- (SEQ. ID NO:139), -FRVN- (SEQ. ID NO:140), -FSIN- (SEQ. ID NO:141), -FKIY- (SEQ. ID NO:142), -FLIN- (SEQ. ID NO:143), -FQIN-(SEQ. ID NO:144), -FQIY-(SEQ. ID NO:145), -FCIN-(SEQ. ID NO:146) , -FCVN- (SEQ. ID NO:147), -FKLN- (SEQ. ID NO:148), -FSVK- (SEQ. ID NO:149), -FSVN- (SEQ. ID NO:150), -FSIK-(SEQ. ID NO:151) , -FYKK- (SEQ. ID NO:152), -FQIK-(SEQ. ID NO:153), -FQIA-(SEQ. ID NO:154) , -FRLQTKY

(SEQ. ID NO:155)- , -FRVYSKY- (SEQ. ID NO:156) and -HRNANT- (SEQ. ID NO:157).

6. The method of claim 1 wherein said defective replication associated protein (Rep) is selected from the group consisting of truncated geminivirus Rep protein, a modified Rep protein capable of binding a geminivirus iteron sequence, or a Rep protein fragment capable of binding a geminivirus iteron sequence.

7. The method of claim 1 wherein introduction of said Rep-iteron antagonist comprises preparing a transgenic plant containing a gene that expresses said replication associated protein or fragment thereof.

8. The method of claim 1 wherein introduction of said Rep-iteron antagonist comprises preparing a transgenic plant having a nucleic acid sequence of a geminivirus iteron capable of binding to a Rep protein.

9. The method of claim 1 wherein introduction of said Rep-iteron antagonist comprises contacting said plant with a composition comprising a nucleic acid molecule selected from the group consisting of an expression vector capable of expressing said replication associated protein and a geminivirus iteron sequence.

10. The method of claim 8 or 9, wherein said contacting comprises biolistic gene transfer or direct DNA uptake into protoplasts.

11. The method of claim 8 or 9, wherein said contacting comprises infection of said plant with a carrier vector.

12. The method of claim 11 wherein said carrier vector is *Agrobacterium*.

13. The method of claim 1 wherein said expression vector is present in a virus particle capable of infecting said plant and expressing said replication associated protein.

14. The method of claim 1 wherein said geminivirus is selected from the group consisting of *Mastrevirus*, *Curtovirus*, *Begomovirus* and *Topcuvirus* genera.

15. The method of claim 14 wherein said *Mastrevirus* species is selected from the group consisting of Bajra streak virus, Bean yellow dwarf virus, Bromus striate mosaic virus, Chickpea chlorotic dwarf virus, Chloris striate mosaic virus, Digitaria streak virus, Digitaria striate mosaic virus, Maize streak virus//Ethiopia, Maize streak virus//Ghana1, Maize streak virus//Ghana2, Maize streak virus//Kenya, Maize streak virus//Komatipoort, Maize streak virus//Malawi, Maize streak virus//Mauritius, Maize streak virus//Mozambique, Maize streak virus//Nigeria1, Maize streak virus//Nigeria2, Maize streak virus//Nigeria3, Maize streak virus//Port Elizabeth, Maize streak virus//Reunion1, Maize streak virus//Reunion2, Maize streak virus//Setaria, Maize streak virus//South Africa, Maize streak virus//Tas, Maize streak virus//Uganda, Maize streak virus//Vaalhart maize, Maize streak virus//Vaalhart wheat, Maize streak virus//Wheat-eleusian, Maize streak virus//Zaire, Maize streak virus//Zimbabwe1, Maize streak virus//Zimbabwe2, Miscanthus streak virus, Panicum streak virus/Karino, Panicum streak virus/Kenya, Paspalum striate mosaic virus, Sugarcane streak virus//Egypt, Sugarcane streak virus/Natal, Sugarcane streak virus/Mauritius, Tobacco yellow dwarf virus, Wheat dwarf virus/Czech Republic (Wheat dwarf virus-CJI, WDV-CJI) Wheat dwarf virus/France and Wheat dwarf virus/Sweden.

16. The method of claim 14 wherein said *Curtovirus* species is selected from the group consisting of Beet curly top virus-California, Beet curly top virus-California//Logan, Beet curly top virus-CFH, Beet curly top virus//Iran, Beet curly top virus-Worland, Horseradish curly top virus and Tomato leafroll virus.

17. The method of claim 14 wherein said *Begomovirus* species is selected from the group consisting of Abutilon mosaic virus, Acalypha yellow mosaic virus, African cassava mosaic virus//Ghana, African cassava mosaic virus/Kenya, African cassava mosaic virus/Nigeria, African cassava mosaic virus/Uganda, Ageratum yellow vein virus, Althea rosea enation virus, Asystasia golden mosaic virus, Bean calico mosaic virus, Bean dwarf mosaic virus, Bean golden mosaic virus-Brazil, Bean golden mosaic virus-Puerto Rico, Bean golden mosaic virus-Puerto Rico/Dominican Rep. (Bean golden mosaic virus-Dominican Rep., BGMV-DR), Bean golden mosaic virus-Puerto Rico/Guatemala (Bean golden mosaic virus-Guatemala, BGMV-GA), Bhendi yellow vein mosaic virus, Chino del tomate virus (Tomato leaf crumple virus, TLCrV), Cotton leaf crumple virus, Cotton leaf curl virus-India, Cotton leaf curl virus-Pakistan1/Faisalabad1 (Cotton leaf curl virus-Pakistan2), Cotton leaf curl virus-Pakistan1/Faisalabad2 (Cotton leaf curl virus-Pakistan3), Cotton leaf curl virus-Pakistan1/Multan (Cotton leaf curl virus-Pakistan1), Cotton leaf curl virus-Pakistan2/Faisalabad (Pakistani cotton leaf curl virus),Cowpea golden mosaic virus, Croton yellow vein mosaic virus//Lucknow, Dolichos yellow mosaic virus, East African cassava mosaic virus/Kenya, East African cassava mosaic virus/Malawi, East African cassava mosaic virus/Tanzania,East African cassava mosaic virus/Uganda//1 (African cassava mosaic virus-Uganda variant), East African cassava mosaic virus/Uganda//2, Eclipta yellow vein virus, Eggplant yellow mosaic virus, Eupatorium yellow vein virus, Euphorbia mosaic virus, Honeysuckle yellow vein mosaic virus, Horsegram yellow mosaic virus, Indian cassava mosaic virus, Jatropha mosaic virus, Leonurus mosaic virus, Limabean golden mosaic virus, Lupin leaf curl virus, Macroptilium golden mosaic virus-Jamaica//2, Macroptilium golden mosaic virus-Jamaica//3, Macrotyloma mosaic virus, Malvaceous chlorosis virus, Melon leaf curl virus, Mungbean yellow mosaic virus, Okra leaf curl virus//Ivory Coast, Okra leaf curl virus//India, Papaya leaf curl virus, Pepper huasteco virus, Pepper golden mosaic virus, (Texas pepper virus), Pepper mild tigrÄ virus, Potato yellow mosaic virus//Guadeloupe, Potato yellow mosaic

virus/Trinidad and Tobago, Potato yellow mosaic virus/Venezuela, Pseuderanthemum yellow vein virus, Rhynchosia mosaic virus, Serrano golden mosaic virus, Sida golden mosaic virus/Costa Rica, Sida golden mosaic virus/Honduras, Sida golden mosaic virus/Honduras//Yellow vein, Sida yellow vein virus, Solanum apical leaf curl virus, Soybean crinkle leaf virus, Squash leaf curl virus, Squash leaf curl virus/Extended host, Squash leaf curl virus/Restricted host, Squash leaf curl virus/Los Mochis, Squash leaf curl virus-China, Tomato golden mosaic virus/Common strain, Tomato golden mosaic virus/Yellow vein strain, Tobacco leaf curl virus//India, Tobacco leaf curl virus-China, Tomato leaf curl virus//Solanum species D1, Tomato leaf curl virus//Solanum species D2, Tomato leaf curl virus-Australia, Tomato leaf curl virus-Bangalore1 (Indian tomato leaf curl virus-BangaloreI), Tomato leaf curl virus-Bangalore2, (Indian tomato leaf curl virus, ItoLCV], Tomato leaf curl virus-Bangalore3 (Indian tomato leaf curl virus- BangaloreII), Tomato leaf curl virus-New Delhi/Severe (Tomato leaf curl virus-India2, ToLCV-IN1), Tomato leaf curl virus-New Delhi/Mild (Tomato leaf curl virus-India2, ToLCV-IN2) Tomato leaf curl virus-New Delhi/Lucknow (Indian tomato leaf curl virus), Tomato leaf curl virus//Senegal, Tomato leaf curl virus-Sinaloa (Sinaloa tomato leaf curl virus, STLCV), Tomato leaf curl virus-Taiwan, Tomato leaf curl virus-Tanzania, Tomato mottle virus, Tomato mottle virus-Taino (Taino tomato mottle virus, TTMoV)], Tomato severe leaf curl virus//Guatemala, Tomato severe leaf curl virus//Honduras, Tomato severe leaf curl virus//Nicaragua, Tomato yellow dwarf virus, Tomato yellow leaf curl virus-China, Tomato yellow leaf curl virus-Israel, Tomato yellow leaf curl virus-Israel/Mild, Tomato yellow leaf curl virus-Israel/Egypt, (Tomato yellow leaf curl virus-Egypt, TYLCV-EG), Tomato yellow leaf curl virus-Israel//Cuba, Tomato yellow leaf curl virus-Israel//Jamaica, Tomato yellow leaf curl virus-Israel//Saudi Arabia1, (Tomato yellow leaf curl virus-Northern Saudi Arabia, TYLCV-NSA), Tomato yellow leaf curl virus-Nigeria, Tomato yellow leaf curl virus-Sardinia, Tomato yellow leaf curl, virus-Sardinia/Sicily (Tomato yellow leaf curl virus-Sicily, TYLCV-SY), Tomato yellow leaf curl virus-Sardinia/Spain//1 (Tomato yellow leaf curl virus-Spain, TYLCV-Sp), Tomato yellow leaf curl

virus-Sardinia/Spain//2 (Tomato yellow leaf curl virus-Almeria, TYLCV-Almeria), Tomato yellow leaf curl virus-Sardinia/Spain//3 (Tomato yellow leaf curl virus-European strain), Tomato yellow leaf curl virus-Saudi Arabia (Tomato yellow leaf curl virus-Southern Saudi Arabia, TYLCV-SSA), Tomato yellow leaf curl virus-Tanzania, Tomato yellow leaf curl virus-Thailand//1, Tomato yellow leaf curl virus-Thailand//2 , Tomato yellow leaf curl virus//Yemen, Tomato yellow mosaic virus-Brazil//1, Tomato yellow mosaic virus-Brazil//2, Tomato yellow mottle virus, Tomato yellow vein streak virus-Brazil, Watermelon chlorotic stunt virus, Watermelon curly mottle virus and Wissadula golden mosaic virus-Jamaica//1.

18. The method of claim 1 wherein said plant is selected from the group consisting of Abutilon, Acalypha, apple, Ageratum, Althea rosea, Asystasia, Bajra, banana, barley, beans, beet, Blackgram, Bromus, Cassava, chickpea, Chilllies, Chloris, clover, coconut, coffee, cotton, cowpea, Croton, cucumber, Digitaria, Dolichos, eggplant, Eupatorium, Euphorbia, fababean, honeysuckle, horsegram, Jatropha, Leonurus, limabean, Lupin, Macroptilium, Macrotyloma, maize, melon, millet, mungbean, oat, okra, Panicum, papaya, Paspalum, peanut, pea, pepper, pigeon pea, pineapple, Phaseolus, potato, Pseuderanthemum, pumpkin, Rhynchosia, rice, Serrano, Sida, sorghum, soybean, squash, sugarcane, sugarbeet, sunflower, sweet potato, tea, tomato, tobacco, watermelon, wheat and Wissadula.

19. The method of claim 1 wherein said replication associated protein (Rep) binds to an iteron sequence selected from the group of sequences selected from GGAGAXGGAGA (SEQ ID NO:99), GGTGTXGGTGT (SEQ ID NO:100) , GGTACXGGTAC (SEQ ID NO:107 ), GGGGAXGGGGA (SEQ ID NO:109 ), GGGGGXGGGGG (SEQ ID NO: 110), GGTGCGCCCXGGCGCACC (SEQ ID NO:101) GCGCCTTCXGAAGGCGCG (SEQ ID NO:102) GGTTTGCGXCGCAAACC (SEQ ID NO:103) GGAGGTGCGTCCX-CCTCCACGGG (SEQ ID NO:105), GGAGTXGGAGT (SEQ ID NO:106 ) and GGTACXGGTAC. (SEQ ID NO:107), GTGAGTGXCACTCAC (SEQ. ID NO:

104), GGTACXGGTAC (SEQ. ID NO:108), GGGGAXGGGGA (SEQ. ID NO:109), GGGGGXGGGGG (SEQ. ID NO:110) wherein "X" is 3-30 nucleotides.

20. The method of claim 1 wherein said replication associated protein (Rep) binds to a DNA sequence comprising GGTGTCTGGAGTC (SEQ ID NO: 111).

21. The method of claim 7 wherein said gene comprises a nucleic acid sequence encoding a polypeptide comprising a sequence selected from the group consisting of -FRVQ- (SEQ. ID NO:130), -FRVN- (SEQ. ID NO:131), -FRIN- (SEQ. ID NO:132), -FRIQ- (SEQ. ID NO:133), -FRLQ- (SEQ. ID NO:134), -FKVQ- (SEQ. ID NO:135), -FKIY-(SEQ. ID NO:136), -FKIN- (SEQ. ID NO:137), -FRLA- (SEQ. ID NO:138), FRLN- (SEQ. ID NO:139), -LKTN- (SEQ. ID NO:140), -FAIN-(SEQ. ID NO:141), -FRLT-(SEQ. ID NO:142), -FNIN- (SEQ. ID NO:143), -FRVN- (SEQ. ID NO:144), -FSIN- (SEQ. ID NO:145), -FKIY- (SEQ. ID NO:146), -FLIN- (SEQ. ID NO:147), -FQIN-(SEQ. ID NO:148), -FQIY-(SEQ. ID NO:149), -FCIN-(SEQ. ID NO:150) , -FCVN- (SEQ. ID NO:151), -FKLN- (SEQ. ID NO:152), -FSVK- (SEQ. ID NO:153), -FSVN- (SEQ. ID NO:154), -FSIK-(SEQ. ID NO:155) , -FYKK- (SEQ. ID NO:156), -FQIK-(SEQ. ID NO:157), -FQIA-(SEQ. ID NO:158) , -FRLQTKY (SEQ. ID NO:159)- , -FRVYSKY- (SEQ. ID NO:160) and -HRNANT- (SEQ. ID NO:161).

22. A vector containing a nucleotide sequence that encodes a defective geminivirus replication associated protein, wherein said encoded protein comprises a polypeptide having an amino acid sequence of a conserved geminivirus iteron binding site or a mutant thereof..

23. The vector of claim 22 wherein said vector is an expression vector that is expressed in plants.

24. The vector of claim 22 wherein said polypeptide comprises an amino acid residue sequence shown in Figure 1

25. The vector of claim 22 wherein said encoded polypeptide forms a dimer with wild type geminivirus Rep protein.

26 . The vector of claim 22 wherein said polypeptide defines from two to thirty different conserved iteron binding sites.

27. The vector according to claim 22 wherein said nucleotide sequence encodes at least two different replication associated proteins.

28. A composition for producing resistance to a geminivirus that infects plants comprising an effective amount of the DNA expression vector of claim 22.

29. A transgenic plant containing a DNA expression vector according to claim 22.

30. An isolated nucleic acid molecule comprising a nucleotide sequence ofa conserved geminivirus iteron.

31. The nucleic acid molecule of claim 30 wherein said nucleotide sequence comprises at least two geminivirus iterons.

32. The nucleic acid molecule according to claim 30 wherein said nucleotide sequence comprises from two to thirty different classes of geminivirus iteron shown in Figure 1A-1C.

33. A composition for producing resistance to a geminivirus that infects plants comprising an effective amount of the nucleic acid molecule of claim 30.

34. A transgenic plant comprising a nucleic acid molecule having a nucleotide sequence comprising a conserved geminivirus iteron.

35. The method of claim 1 wherein said antagonist is the nucleic acid of claim 30.

36. An isolated DNA sequence comprising GGTGTCTGGAGTC (SEQ ID NO: ).

37. The progeny of the transgenic plant of claim 29 or 34, wherein said progeny expresses a defective Rep protein or a conserved geminivirus iteron.

38. A seed of the transgenic plant of claim 29 or 34.

39. An isolated polypeptide selected from the group consisting of $AC1_{1-21}$, $AC1_{1-60}$, $AC1_{1-52}$, $AC1_{1-180}$, $AC1_{1-111}$, $AC1_{1-114}$ and $AC1_{1-160}$.

40. An isolated nucleic acid comprising a sequence encoding a protein of claim 39.

41. A truncated Rep protein.

42. The truncated protein of claim 41 wherein said protein comprises least $AC1_{1-160}$.

43. A method for inhibiting geminivirus replication in a plant comprising introducing a geminivirus replication associated protein (Rep)-iteron antagonist into said plant, said antagonist selected from the group consisting of a nucleotide sequence defining a geminivirus iteron capable of binding to a Rep protein and a defective Rep, wherein said defective Rep comprises a conserved geminivirus iteron binding site.

44. A method for providing resistance to infection by geminiviruses in a susceptible plant comprising:

a) transforming susceptible plant cells with a DNA molecule that comprises operatively linked in sequence in the 5' to 3' direction:

i) a promoter region that functions in plant cells to cause the production of an RNA sequence; and

ii) a gene encoding a defective Rep protein, wherein said defective Rep comprises a conserved geminivirus iteron binding site;

b) selecting said plant cells that have been transformed;

c) regenerating said plant cells to provide a differentiated plant; and

d) selecting a transformed plant that expresses said defective Rep gene at a level sufficient to render the plant at least partially resistant to infection by the geminivirus.

45. An at least partially virus-resistant transformed plant normally susceptible to infection by a geminivirus having inserted into its genome a DNA molecule that comprises operatively linked in sequence in the 5' to 3' direction:

i) a promoter region that functions in plant cells to cause the production of an RNA sequence; and

ii) a gene encoding a defective Rep protein, wherein said defective Rep comprises a conserved geminivirus iteron binding site.

46. The truncated Rep protein of claim 41 which comprises an amino acid sequence corresponding to amino acids 1- 160 wherein said truncated Rep protein inhibits at least partially, geminivirus infection or geminivirus replication in a transgenic plant.

47. The method of claim 46, wherein said truncated Rep protein at least partially inhibits geminivirus infection or geminivirus replication in *Nicotiana*.

48. The method of claim 14 wherein said topcuvirus species is Tomato pseudo-curly top virus.

49. The nucleic acid molecule according to claim 30 wherein said nucleotide sequence comprises from two to thirty of the same classes of geminivirus iteron shown in Figure 1A-1C.

50. The truncated Rep protein of claim 41 which comprises an amino acid sequence corresponding to amino acids 1-52 wherein said truncated Rep protein inhibits at least partially, geminivirus infection or geminivirus replication in a transgenic plant.

51. A method for producing at least partial resistance to a virus or a method for reducing replication of a virus in a plant, plant cell, propagule, animal or animal cell, comprising introducing a replication associated protein (Rep)-iteron antagonist into a plant, plant cell, propagule, animal or animal cell, wherein said antagonist is selected from the group consisting of a nucleotide sequence of an iteron capable of binding to a Rep protein and a defective Rep, wherein said defective Rep comprises a conserved iteron binding site, and wherein said Rep-iteron antagonist renders the infected plant, plant cell, propagule, animal or animal cell at least partially resistant to the infection.

52. The method of claim 51 wherein the Rep-iteron antagonist sequence is at least 50% identical to those found in Figs 1A-1C (SEQ ID NOS:1-8, 10-35, 37-41, 43-49, 54-57, 59-62, 64-107).

53. The method of claim 51 wherein the Rep-iteron antagonist sequence is at least 60%, 70%, 80%, 90% 95% or 99% identical to those found in Figs 1A-1C (SEQ ID NOS:1-8, 10-35, 37-41, 43-49, 54-57, 59-62, 64-107).

54. The method of claim 51 wherein the viral infection is from a *Nanovirus or Circoviridae.*

55. The method of claim 51 wherein said virus is a virus that replicates in a manner similar to the geminivirus, i.e. dependent on the binding of a Rep protein to an iteron.

56. A composition for producing at least partial resistance to a virus or for reducing replication of a virus in a plant, plant cell, propagule, animal or animal cell wherein said composition is used in any of the methods of claims 51-55.

57. A Rep-iteron antagonist comprising a nucleic acid sequence encoding a protein that binds to an iteron wherein viral infection or DNA replication of the virus causing the infection is reduced following said antagonist binding to an iteron..

58. A Rep-iteron antagonist comprising a nucleic acid sequence that competes for binding of a Rep protein with the iteron of a virus causing the infection, wherein viral infection or DNA replication of the virus causing the infection is reduced following said antagonist binding to the Rep protein.

59. A Rep-iteron antagonist comprising a polypeptide or the nucleic acid sequence encoding a polypeptide comprising the sequence FLTY or KAYIDK.

60. A Rep-iteron antagonist comprising a polypeptide or the nucleic acid sequence encoding a polypeptide selected from the group consisting of FLTYPqC wherein q is a basic or a polar amino acid, HIHxUUQ wherein U is a bulky hyrophobic residue and xxYxxK wherein x may be any amino acid.

61. A vector comprising a nucleic acid sequence encoding any of the Rep-iteron antagonists of claims 57-60.

## BEGOMOVIRUSES

| Virus | Rep N-terminal sequence | SEQ ID NO: | Iteron |
|---|---|---|---|
| AbMV (X15983) | MPP-PKKFRVQAKNYFLTYP | 1 | GGAGA---GGAGA 99 |
| ACMV/KE (J02057) | MRT-P-RFRIQAKNVFLTYP | 2 | |
| ACMV/NG (X17045) | MRT-P-RFRVQAKNVFLTYP | 3 | |
| ACMV/UG (Z83253) | MRT-P-RFRVQAKNVFLTYP | 4 | |
| BDMV (M88179) | MPP-PKKFRVQSRNYFLTYP | 5 | |
| BGMV-PR (M10070) | MPP-PQRFRVQSKNYFLTYP | 6 | |
| BGMV-PR/DO (L01635) | MPP-PKRFRINAKNYFLTYP | 7 | |
| BGMV-PR/GA (M91604) | MPP-PQRFRVQSKNYFLTYP | 8 | |
| CLCuV-PK/26 (AJ002447) | MAP-PKRFKVQAKNYFITYP | 10 | |
| CLCuV-PK/62 (AJ002448) | MAP-PKRFKVQAKNYFITYP | 11 | |
| ClCuV-PK1/F (AJ002452) | MAP-PKRFKVQAKNYFITYP | 12 | |
| LMV-BR (U92532) | MPSKPRRFRVQAKNIFLTYP | 13 | |
| SiGMV-CR (X99550) | MPP-PKKFRVQSKNHFLTYP | 14 | |
| SiGMV-Flo (UT6964) | MPP-PKKFRVQSKNHFLTYP | 15 | |
| SiGMV-HO (X11097) | MPP-PKRFRVNSKNYFLTYP | 16 | |
| SiGMV-HO/Yv (Y11099) | MPP-PKKFRVQSKNHFLTYP | 17 | |
| ToMoV (L14460) | MPP-PKRFRVNSKNYFLTYP | 18 | |
| ToMoV-Maxwell (M90494 M90494) | MPP-PKKFRVQSKNYFLTYP | 19 | |
| WACMV/CI (AF112352) | MRT-P-RFRIQAKNVFLTYP | 20 | |
| AREV | MAP-PKRFKIYAKNYFLTYP | 21 | GGTGT---GGTGT 100 |
| BGMV/BR (M88686) | MPP-PKRFKINAKNYFLTYP | 22 | |
| CaLCuV (U65529) | MPRNPKSFRLAARNIFLTHL | 23 | |
| CLCuV-PK/OK (AJ002459) | M---PRALKTNAKNYFLTYP | 24 | |
| CPGMV (AF029217) | MPR-SGAFRVNAKNIFATYP | 25 | |
| MYMV (Q14203) | MPR-LGRFAINAKNYFLTYP | 26 | |
| PHV (X70418) | MPL-PKRFRLNAKNYFLTYP | 27 | |
| PYMV/TT (AF039031) | MPP--KRFRINSKNYFLTYP | 28 | |
| SqLCV (M38182) | MPRNPNSFRLTARNIFLTYP | 29 | |
| TbLCV/CN (S77090) | MAP-PNKFRINAKNYFLTYP | 30 | |
| ToLCV-AU (S53251) | MTR-PKSFRINAKNYFLTYP | 31 | |
| ToLCV-Ban2 (Z48182) | MPR-PGRFNINAKNYFLTYP | 32 | |
| ToLCV-Nde/s (U15015) | MAS-PRRFRVNAKNYFLTYP | 33 | |
| TOLCV-PA (Y15034) | MPR-AGRFSINARNYFLTYP | 34 | |
| TOLCV-TW (U88692) | MHP-LNKFRINAKNYFLTYP | 35 | |
| TYLCV-IL (X15656) | M---PRLFKIYAKNYFLTYP | 37 | |
| TYLCV-IL/CU (AJ220535) | M---PRLFKIYAKNYFLTYP | 38 | |
| TYLCV-IL/EG (L12249/Mugla) | M---PRLFKIYAKNYFLTYP | 39 | |
| TYLCV-TH//1 (X63015) | MPP-SKKFLINAKNYFLTYP | 40 | |
| TYLCV-TH//2 (AF141897) | MPP-SKKFLINAKNYFLTYP | 41 | |

The spacing between the iterons of different viruses may vary (3-30 nucleotides) even though they have the same sequence.

FIG. 1A

## MASTREVIRUSES

| Virus | Rep-N-terminal sequences | Seq ID NOS: | Iteron* |
|---|---|---|---|
| DSV (M23022) | ----NRSFRHRNANT----------FLTYS 43 | GGTGCGCCC---GGGCGCACC 101 | |
| MSV-N2A/R2 | ---SNRQFSHRNANT----------FLTYP 44 | GCGCCTTC---GAAGGCGCG 102 (AJ224504) | |
| MSV-N2A/R3 | ---SNRQFSHRNANT----------FLTYP 45 | (AJ224505) | |
| MSV-N2A/R4 | ---SNRQFSHRNANT----------FLTYP 46 | (AJ224506) | |
| MSV-N2A/R5 | ---SNRQFSHRNANT----------FLTYP 47 | (AJ224507) | |
| MSV-N2A/R6 | ---SNRQFSHRNANT----------FLTYP 48 | (AJ224508) | |
| MSV-N2A/R8 | ---SNRQFSHRNANT----------FLTYP 49 | (AJ225006) | |
| MSV-SP2/14 | ---SNRQFSHRNANT----------FLTYP 54 | (AJ225011) | |
| MSV-SP2/R10 | ---SNRQFSHRNANT----------FLTYP 55 | (AJ225009) | |
| MSV-SP2/R11 | ---SNRQFSHRNANT----------FLTYP 56 | (AJ225008) | |
| MSV-SP2R12 | ---SNRQFSHRNANT----------FLTYP 57 | (AJ225010) | |
| MSV//KE | ---SNRQFSHRNANT----------FLTYP 60 | (X01085) | |
| MSV//NG | ---SNRQFSHRNANT----------FLTYP 60 | (X02026) | |
| MSV//RE | ---SNRQFSHRNANT----------FLTYP 61 | (X94330) | |
| MSV//RE2 | ---SNRQFSHRNANT----------FLTYP 62 | (X01633) | |
| PanSV/Kar | --HSVRSFRHRNANT----------FLTYS 64 | (L37638) | |
| PanSV/Nat | --HSVRSFRHRNANT----------FLTYS 63 | (X60168) | |
| SSV/Nat | ---PSRRFKHRNVNT----------FLTYS 61 | GGTTTGCG---CGCAAACC 103 (S64567) | |
| TYDV | T----------K$FRLQTKKVFLTYP 64 | GTGAGTG---CACTCAC 104 (M81103) | |
| WDV/CZ | ------PR----------FRVYSKKLFLTYP 68 | (X02865) GGAGGTGCGTCC---CCTCCACGGG 105 | |
| WDV/SW | ------PR----------FRVYSKKLFLTYP 64 | (Q00305) | |

## CURTOVIRUSES

| | | | |
|---|---|---|---|
| BCTV-CAL | MPPT-KRFRIQAKNIFLTYPQ 70 | (X04144) GGAGT---GGAGT 106 | |
| BCTV-CAL/L | MPRTPKRFRIQAKNIFLTYPQ 71 | (M24559) | |
| BCTV-CFH | MP-----FYKKAKNFFLTYPQ 72 | (U02311) | |
| BCTV-WOR | MPRSPS-FQIKAKNIFLTYPR 73 | (U56935) | |
| HrCTV | MPRQPNSFRIQARNIFLTYPQ 74 | (U49907) | |
| TOPCUVIRUS TPCTV | MPSNPKRFQIAAAKNYLTYPN 75 | (X84735) GGTAC---GGTAC 107 | |

* The spacing between the iterons of different viruses may vary (3-30 bases) even though they have the same sequence.

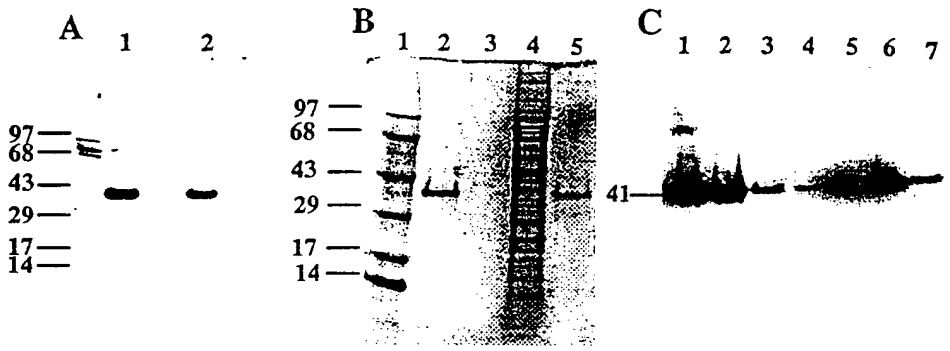
FIG. 1B

## BEGOMOVIRUSES

| Virus | Rep-N-terminal sequences | | Iteron* | |
|---|---|---|---|---|
| | | Seq ID nos. | | |
| AToLCV | MPP-PKRFQINSKNYFLTYP | 76 | GGTAC---GGTAC | 108 |
| AYVV (X74516) | MAP-PKRFQINAKNYFLTYP | 77 | | |
| EACMV/TZ (Z83256) | MPR-AGRFQINAKNYFITYP | 78 | | |
| EACMV-UG//1 (U12534) | MPR-AGRFQINAKNYFVTYP | 79 | | |
| EACMV-UG//2 (Z83257) | MPR-AGRFQINAKNYFITYP | 80 | | |
| ICMV (Z24758) | MSP-PKRFQINAKNYFLTYP | 81 | | |
| OYVMV/201 (AJ002451) | MAP-PKQFQIYAKNYFITYP | 82 | | |
| TGMV/Com (M73794) | M-P-PKRFLINSKNYFLTYP | 83 | | |
| TYLCV-IL/Mld | MPSHPKRFQINAKNYFLTYP | 84 | | |
| TYLCV-Sar/E (X61153) | MAP-PKRFQINCKNYFLTYP | 85 | | |
| TYLCV-Sar/E (X61153) | MAQ-PKRFQINAKHYFLSFP | 86 | | |
| TYLCV-Sar/S (Z28390) | MAQ-PKRFQINAKHYFLTFP | 87 | | |
| EACMV/CI (X76319) | MPR-AGRFQINAKNYFITYP | 88 | | |
| ClCuV-PK80 (AJ222703) | MPR-NNSFCINAKNIFLTFP | 89 | GGGGA---GGGGA | 109 |
| ClCuV-PK/80 (AJ002455) | MPR-NNSFCINAKNIFLTFP | 90 | | |
| PaLCV (X07962) | MPR-LNSFCVNAKNIFLTYP | 91 | | |
| ToLCV-Ban3 (U38239) | MAA-PNRFKLNAKNYFLTYP | 92 | | |
| ToMov-Tai (U88692) | MPR-KGSFSVKAKNYFLTYP | 93 | | |
| ToYVSV (U79998) | MPP-PKRFSVNAKNFFLTYP | 94 | | |
| PYMV-VE (U00940) | MPRKGS-FSIKAKNYFLTYP | 95 | GGGGG---GGGGG | 110 |
| PYMV-VE/To (AF039031) | MPRKGS-FSIKAKNYFLTYP | 96 | | |
| ToDLCV-JM (V84146) | MPR-KGYFSVKAKNYFLTYP | 97 | | |
| TYLCV-Sar (Z28390) | MPR-SGRFSIKAKNYFLTYP | 98 | | |

*The spacing between the iterons of different viruses may vary (3-30 nucleotides) even though they have the same sequence.

Fig. 1C




Figs 2A-2C

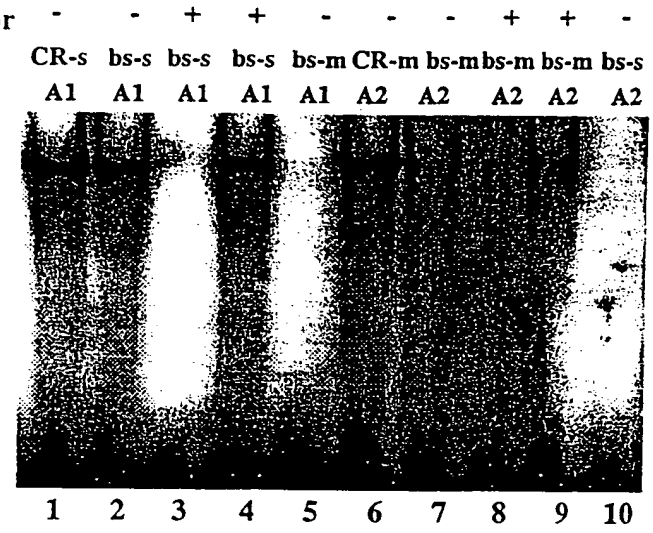
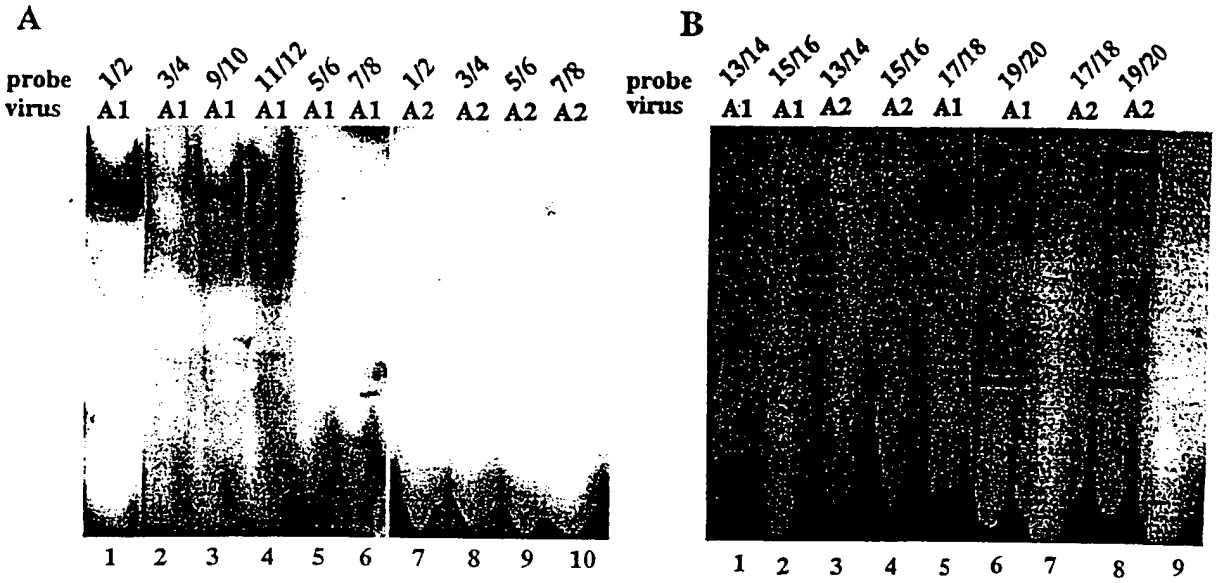

competitior - - + + - - - + + -
probe CR-s bs-s bs-s bs-s bs-m CR-m bs-m bs-m bs-m bs-s
virus A1 A1 A1 A1 A1 A2 A2 A2 A2 A2
1 2 3 4 5 6 7 8 9 10

Fig 3


A
probe
virus
1/2 A1
3/4 A1
9/10 A1
11/12 A1
5/6 A1
7/8 A1
1/2 A2
3/4 A2
5/6 A2
7/8 A2
1 2 3 4 5 6 7 8 9 10
B
probe
virus
13/14 A1
15/16 A1
13/14 A2
15/16 A2
17/18 A1
19/20 A1
17/18 A2
19/20 A2
1 2 3 4 5 6 7 8 9

Figs. 4A-4B

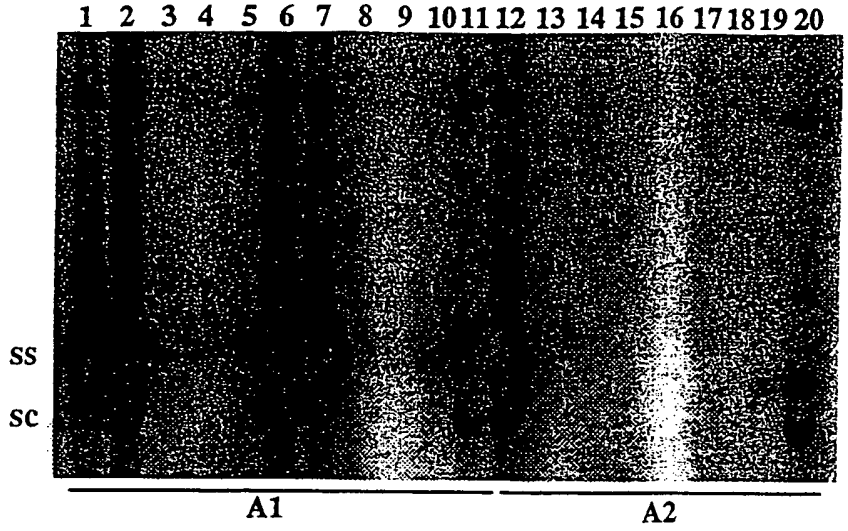
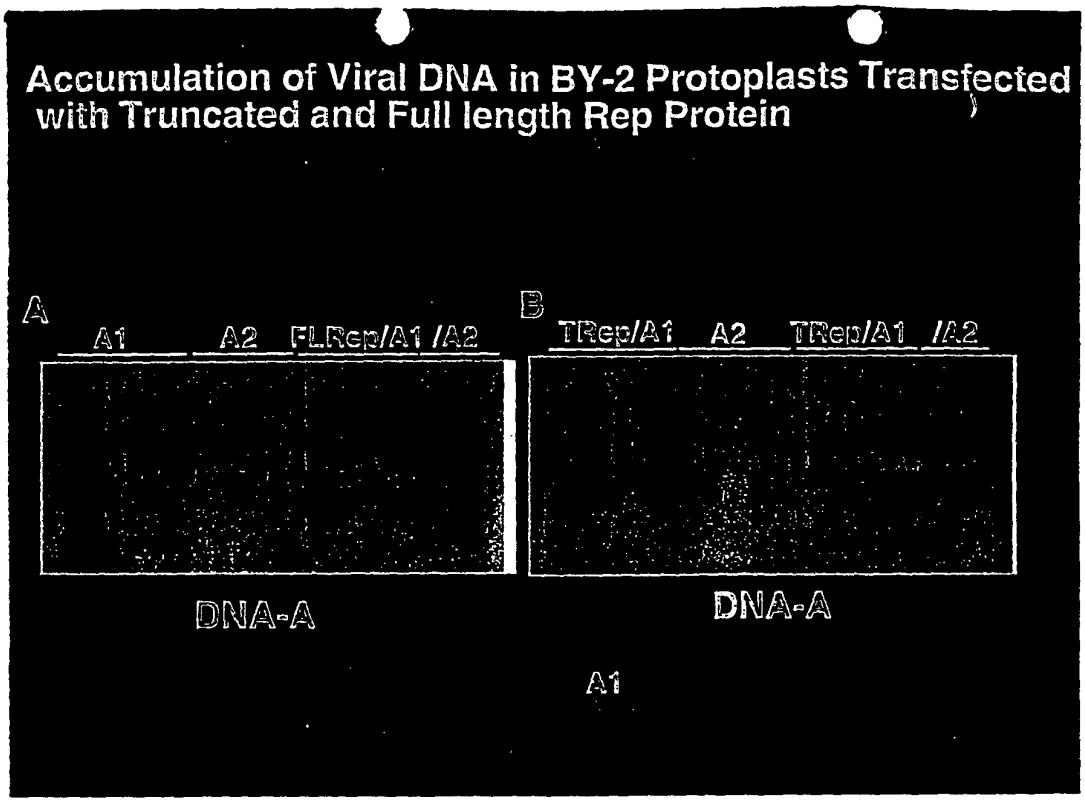



Fig 5


Accumulation of Viral DNA in BY-2 Protoplasts Transfected with Truncated and Full length Rep Protein
A
A1
A2
FLRep/A1
/A2
DNA-A
B
TRep/A1
A2
TRep/A1
/A2
DNA-A
A1

Figs 6A-6B

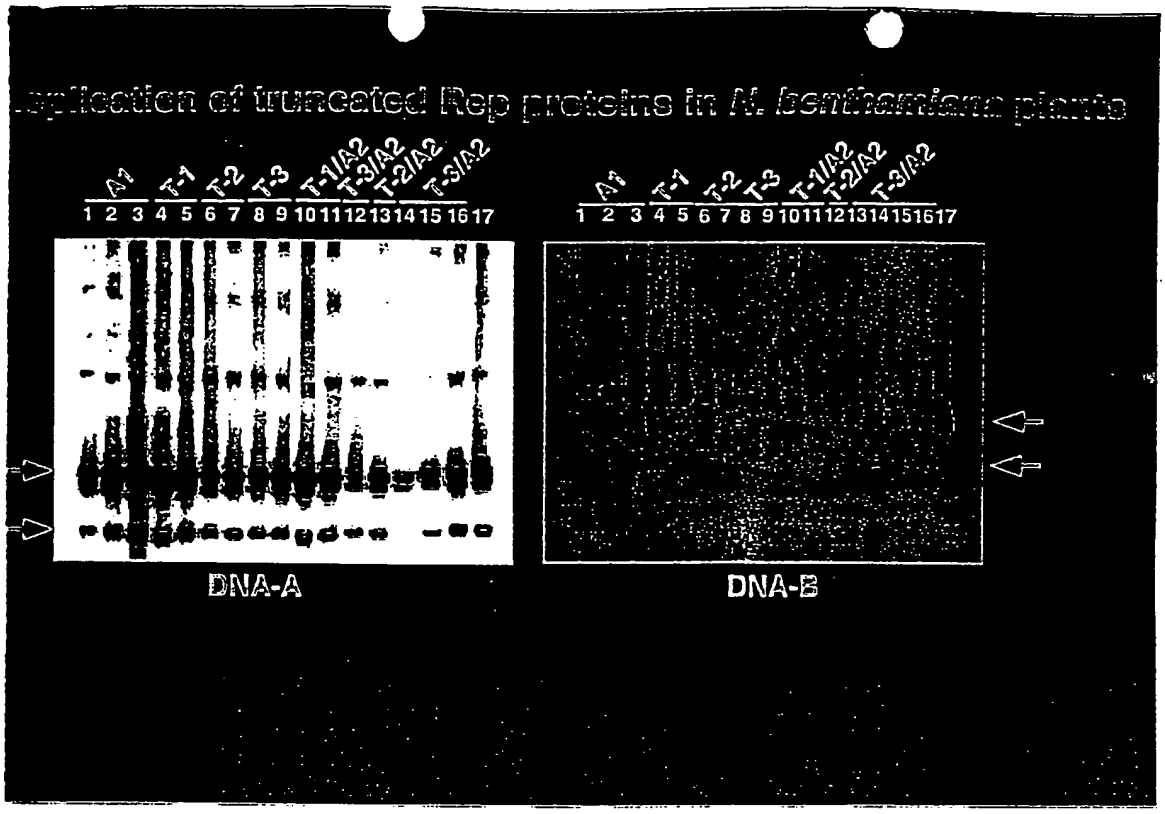
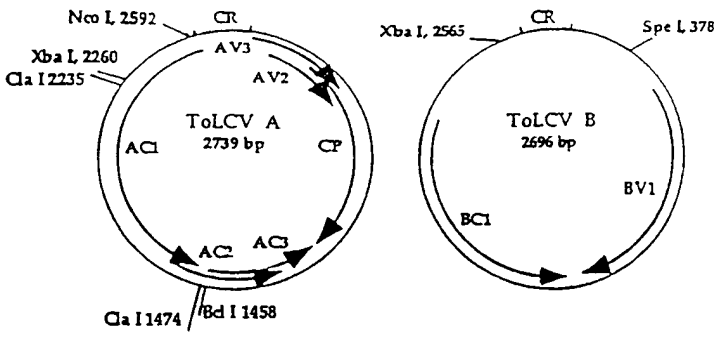
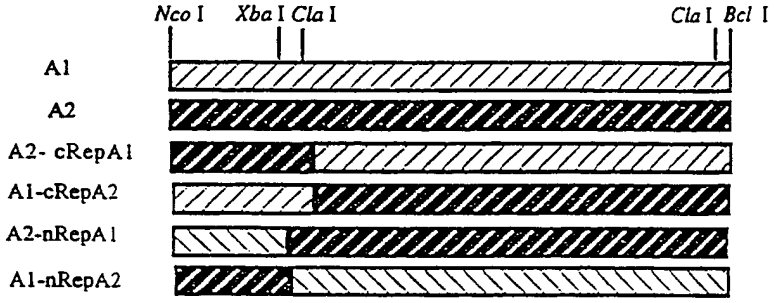

Replication of truncated Rep proteins in N. benthamiana plants
A1 T-1 T-2 T-3 T-1/A2 T-3/A2 T-2/A2 T-3/A2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
DNA-A
A1 T-1 T-2 T-3 T-1/A2 T-2/A2 T-3/A2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
DNA-B

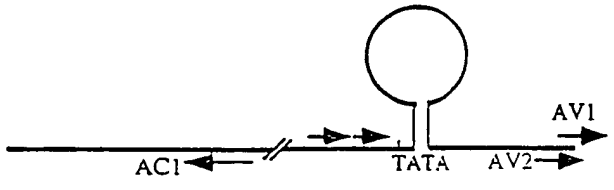
Fig 7




B




C




| SEQ ID NO: | | | | | | | | | | | | | | | | SEQ ID NO: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 112 | A | N | V | R | F | R | R | P | P | A | M | GGTGTCTGGAGTC | A1 | | | 121 |
| 113 | A | D | I | R | F | R | R | P | P | A | M | GGCGTCTGGCGTC | A2 | | | 122 |
| 114 | A | N | V | R | F | R | R | P | P | A | M | GGTGTCTGGAGTC | A2-Repm1/CRm3 | | | 123 |
| 115 | A | N | I | R | F | R | R | P | P | A | M | GGTGTCTGGAGTC | A2Repm2/CRm3 | | | 124 |
| 116 | A | N | V | R | F | R | R | P | P | A | M | GG-GTCTGGAGTC | A1-CRm1 | | | 125 |
| 117 | A | N | V | R | F | R | R | P | P | A | M | GGTGTCTGG-GTC | A1-CRm2 | | | 126 |
| 118 | A | D | V | R | F | R | R | P | P | A | M | GGTGTCTGGAGTC | A1-Repm3 | | | 127 |
| 119 | A | N | V | R | F | R | R | P | P | A | M | GGCGTCTGGAGTC | A1-CRm4 | | | 128 |
| 120 | A | D | V | R | F | R | R | P | P | A | M | GGCGTCTGGAGTC | A1-Repm4/Crm4 | | | 129 |

FIG. 8A-8C

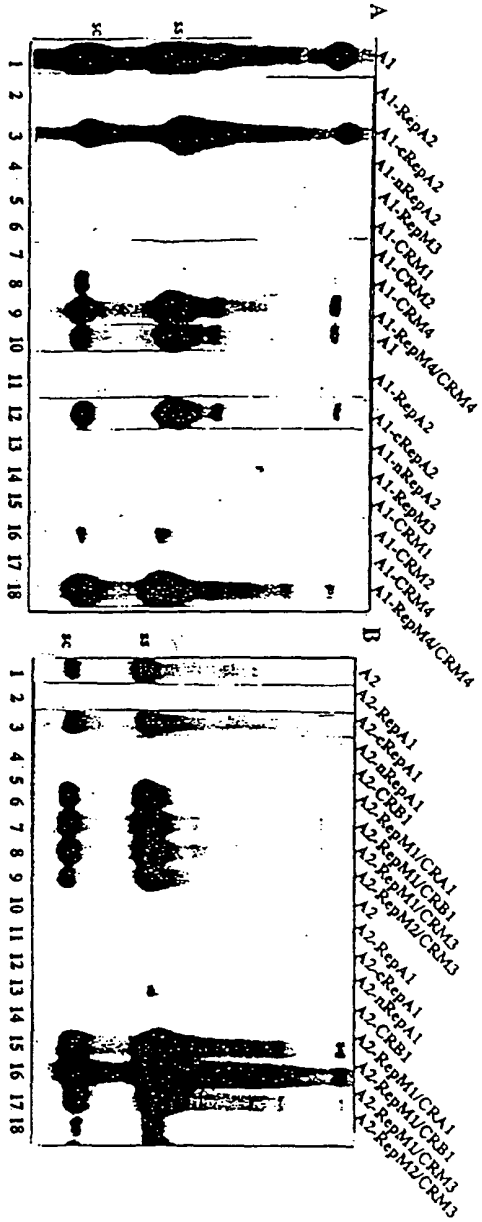

A
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
A1
A1-RepA2
A1-cRepA2
A1-nRepA2
A1-RepM3
A1-CRM1
A1-CRM2
A1-CRM4
A1-RepM4/CRM4
A2
A2-RepA1
A2-cRepA1
A2-nRepA1
A2-CRB1
A2-RepM1/CRA1
A2-RepM1/CRB1
A2-RepM1/CRM3
A2-RepM2/CRM3

FIG. 9.

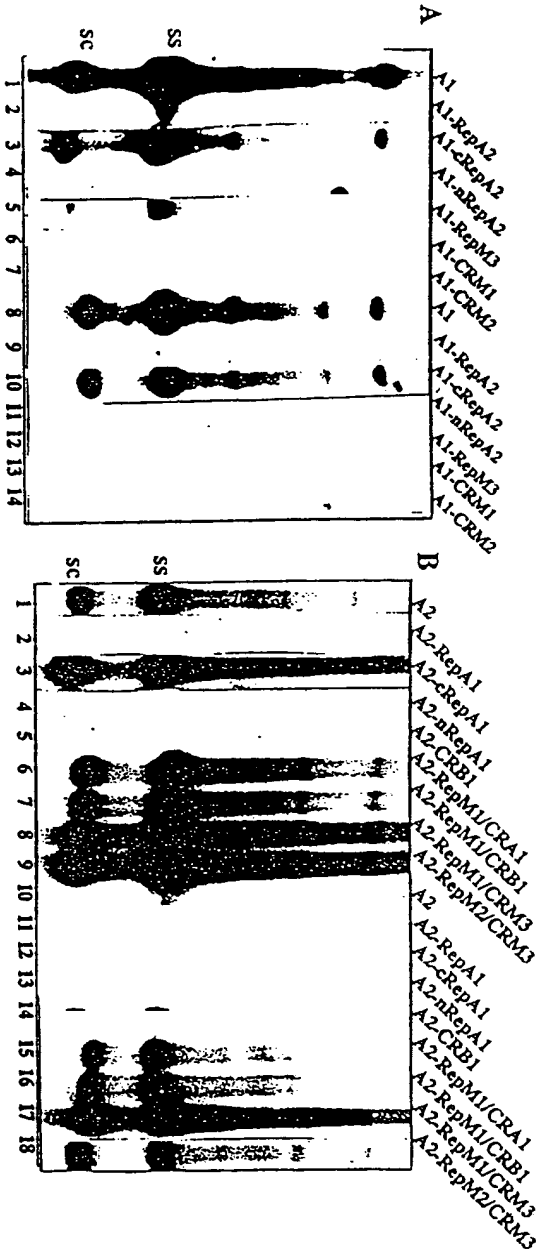

A
sc
ss
1 2 3 4 5 6 7 8 9 10 11 12 13 14
A1
A1-RepA2
A1-cRepA2
A1-nRepA2
A1-RepM3
A1-CRM1
A1-CRM2
A1
A1-RepA2
A1-cRepA2
A1-nRepA2
A1-RepM3
A1-CRM1
A1-CRM2
B
sc
ss
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
A2
A2-RepA1
A2-cRepA1
A2-nRepA1
A2-CRB1
A2-RepM1/CRA1
A2-RepM1/CRB1
A2-RepM1/CRM3
A2-RepM2/CRM3
A2
A2-RepA1
A2-cRepA1
A2-nRepA1
A2-CRB1
A2-RepM1/CRA1
A2-RepM1/CRB1
A2-RepM1/CRM3
A2-RepM2/CRM3

FIG. 10

SEQUENCE LISTING

<110> The Scripps Research Institute *et al.*

<120> Control of Virus Infection Using Replication Associated Proteins, Compositions and Methods of Use

<130> 1294.002PC02

<140>
<141>

<150> 60/117,285
<151> 1999-01-26

<160> 208

<170> PatentIn Ver. 2.1

<210> 1
<211> 19
<212> PRT
<213> Begomovirus

<400> 1
Met Pro Pro Pro Lys Lys Phe Arg Val Gln Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 2
<211> 18
<212> PRT
<213> Begomovirus

<400> 2
Met Arg Thr Pro Arg Phe Arg Ile Gln Ala Lys Asn Val Phe Leu Thr
1 5 10 15

Tyr Pro

<210> 3
<211> 18
<212> PRT
<213> Begomovirus

<400> 3
Met Arg Thr Pro Arg Phe Arg Val Gln Ala Lys Asn Val Phe Leu Thr
1 5 10 15

Tyr Pro

<210> 4

<210> 4
<211> 18
<212> PRT
<213> Begomovirus

<400> 4
Met Arg Thr Pro Arg Phe Arg Val Gln Ala Lys Asn Val Phe Leu Thr
1 5 10 15

Tyr Pro

<210> 5
<211> 19
<212> PRT
<213> Begomovirus

<400> 5
Met Pro Pro Pro Lys Lys Phe Arg Val Gln Ser Arg Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 6
<211> 19
<212> PRT
<213> Begomovirus

<400> 6
Met Pro Pro Pro Gln Arg Phe Arg Val Gln Ser Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 7
<211> 19
<212> PRT
<213> Begomovirus

<400> 7
Met Pro Pro Pro Lys Arg Phe Arg Ile Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 8
<211> 19
<212> PRT
<213> Begomovirus

<400> 8
Met Pro Pro Pro Gln Arg Phe Arg Val Gln Ser Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 9
<211> 15
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (8)
<223> x can be 3-30 nucleotides

<400> 9
gtgagtgxca ctcac 15

<210> 10
<211> 19
<212> PRT
<213> Begomovirus

<400> 10
Met Ala Pro Pro Lys Arg Phe Lys Val Gln Ala Lys Asn Tyr Phe Ile
1 5 10 15

Thr Tyr Pro

<210> 11
<211> 19
<212> PRT
<213> Begomovirus

<400> 11
Met Ala Pro Pro Lys Arg Phe Lys Val Gln Ala Lys Asn Tyr Phe Ile
1 5 10 15

Thr Tyr Pro

<210> 12
<211> 19
<212> PRT
<213> Begomovirus

<400> 12
Met Ala Pro Pro Lys Arg Phe Lys Val Gln Ala Lys Asn Tyr Phe Ile
1 5 10 15

Thr Tyr Pro

<210> 13
<211> 20
<212> PRT
<213> Begomovirus

<400> 13

Met Pro Ser Lys Pro Arg Arg Phe Arg Val Gln Ala Lys Asn Ile Phe
1 5 10 15

Leu Thr Tyr Pro
20

<210> 14
<211> 19
<212> PRT
<213> Begomovirus

<400> 14
Met Pro Pro Pro Lys Lys Phe Arg Val Gln Ser Lys Asn His Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 15
<211> 19
<212> PRT
<213> Begomovirus

<400> 15
Met Pro Pro Pro Lys Lys Phe Arg Val Gln Ser Lys Asn His Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 16
<211> 19
<212> PRT
<213> Begomovirus

<400> 16
Met Pro Pro Pro Lys Arg Phe Arg Val Asn Ser Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 17
<211> 19
<212> PRT
<213> Begomovirus

<400> 17
Met Pro Pro Pro Lys Lys Phe Arg Val Gln Ser Lys Asn His Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 18
<211> 19
<212> PRT

<213> Begomovirus

<400> 18
Met Pro Pro Pro Lys Arg Phe Arg Val Asn Ser Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 19
<211> 19
<212> PRT
<213> Begomovirus

<400> 19
Met Pro Pro Pro Lys Lys Phe Arg Val Gln Ser Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 20
<211> 18
<212> PRT
<213> Begomovirus

<400> 20
Met Arg Thr Pro Arg Phe Arg Ile Gln Ala Lys Asn Val Phe Leu Thr
1 5 10 15

Tyr Pro

<210> 21
<211> 19
<212> PRT
<213> Begomovirus

<400> 21
Met Ala Pro Pro Lys Arg Phe Lys Ile Tyr Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 22
<211> 19
<212> PRT
<213> Begomovirus

<400> 22
Met Pro Pro Pro Lys Arg Phe Lys Ile Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 23
<211> 20
<212> PRT
<213> Begomovirus

<400> 23
Met Pro Arg Asn Pro Lys Ser Phe Arg Leu Ala Ala Arg Asn Ile Phe
1 5 10 15

Leu Thr His Leu
20

<210> 24
<211> 17
<212> PRT
<213> Begomovirus

<400> 24
Met Pro Arg Ala Leu Lys Thr Asn Ala Lys Asn Tyr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 25
<211> 19
<212> PRT
<213> Begomovirus

<400> 25
Met Pro Arg Ser Gly Ala Phe Arg Val Asn Ala Lys Asn Ile Phe Ala
1 5 10 15

Thr Tyr Pro

<210> 26
<211> 19
<212> PRT
<213> Begomovirus

<400> 26
Met Pro Arg Leu Gly Arg Phe Ala Ile Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 27
<211> 19
<212> PRT
<213> Begomovirus

<400> 27
Met Pro Leu Pro Lys Arg Phe Arg Leu Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 28
<211> 18
<212> PRT
<213> Begomovirus

<400> 28
Met Pro Pro Lys Arg Phe Arg Ile Asn Ser Lys Asn Tyr Phe Leu Thr
1 5 10 15

Tyr Pro

<210> 29
<211> 20
<212> PRT
<213> Begomovirus

<400> 29
Met Pro Arg Asn Pro Asn Ser Phe Arg Leu Thr Ala Arg Asn Ile Phe
1 5 10 15

Leu Thr Tyr Pro
20

<210> 30
<211> 19
<212> PRT
<213> Begomovirus

<400> 30
Met Ala Pro Pro Asn Lys Phe Arg Ile Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 31
<211> 19
<212> PRT
<213> Begomovirus

<400> 31
Met Thr Arg Pro Lys Ser Phe Arg Ile Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 32
<211> 19
<212> PRT
<213> Begomovirus

<400> 32
Met Pro Arg Pro Gly Arg Phe Asn Ile Asn Ala Lys Asn Tyr Phe Leu

1 5 10 15

Thr Tyr Pro

<210> 33
<211> 19
<212> PRT
<213> Begomovirus

<400> 33
Met Ala Ser Pro Arg Arg Phe Arg Val Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 34
<211> 19
<212> PRT
<213> Begomovirus

<400> 34
Met Pro Arg Ala Gly Arg Phe Ser Ile Asn Ala Arg Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 35
<211> 19
<212> PRT
<213> Begomovirus

<400> 35
Met His Pro Leu Asn Lys Phe Arg Ile Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 36
<211> 11
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (6)
<223> x can be 3-30 nucleotides

<400> 36
ggtacxggta c 11

<210> 37
<211> 17
<212> PRT

<213> Begomovirus

<400> 37
Met Pro Arg Leu Phe Lys Ile Tyr Ala Lys Asn Tyr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 38
<211> 17
<212> PRT
<213> Begomovirus

<400> 38
Met Pro Arg Leu Phe Lys Ile Tyr Ala Lys Asn Tyr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 39
<211> 17
<212> PRT
<213> Begomovirus

<400> 39
Met Pro Arg Leu Phe Lys Ile Tyr Ala Lys Asn Tyr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 40
<211> 19
<212> PRT
<213> Begomovirus

<400> 40
Met Pro Pro Ser Lys Lys Phe Leu Ile Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 41
<211> 19
<212> PRT
<213> Begomovirus

<400> 41
Met Pro Pro Ser Lys Lys Phe Leu Ile Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 42
<211> 11
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (6)
<223> x can be 3-30 nucleotides

<400> 42
ggggaxgggg a 11

<210> 43
<211> 16
<212> PRT
<213> Mastrevirus

<400> 43
Asn Arg Ser Phe Arg His Arg Asn Ala Asn Thr Phe Leu Thr Tyr Ser
1 5 10 15

<210> 44
<211> 17
<212> PRT
<213> Mastrevirus

<400> 44
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 45
<211> 17
<212> PRT
<213> Mastrevirus

<400> 45
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 46
<211> 17
<212> PRT
<213> Mastrevirus

<400> 46
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 47
<211> 17
<212> PRT
<213> Mastrevirus

<400> 47
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 48
<211> 17
<212> PRT
<213> Mastrevirus

<400> 48
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 49
<211> 17
<212> PRT
<213> Mastrevirus

<400> 49
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro 11

<210> 50
<211> 11
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (6)
<223> x can be 3-30 nucleotides

<400> 50
gggggxgggg g

<210> 51
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 51
ggggtctgga gtc 13

<210> 52
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 52
ggggtctgga gtc 13

<210> 53
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 53
ggtgtctggt gtc 13

<210> 54
<211> 17
<212> PRT
<213> Mastrevirus

<400> 54
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 55
<211> 17
<212> PRT
<213> Mastrevirus

<400> 55
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 56
<211> 17
<212> PRT

<213> Mastrevirus

<400> 56
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 57
<211> 17
<212> PRT
<213> Mastrevirus

<400> 57
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 58
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 58
ggagtctgga gtc 13

<210> 59
<211> 17
<212> PRT
<213> Mastrevirus

<400> 59
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 60
<211> 17
<212> PRT
<213> Mastrevirus

<400> 60
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 61
<211> 17
<212> PRT
<213> Mastrevirus

<400> 61
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 62
<211> 17
<212> PRT
<213> Mastrevirus

<400> 62
Ser Asn Arg Gln Phe Ser His Arg Asn Ala Asn Thr Phe Leu Thr Tyr
1 5 10 15

Pro

<210> 63
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synhetic oligonucleotide

<400> 63
ggtgtctttt ttggagtc 18

<210> 64
<211> 18
<212> PRT
<213> Mastrevirus

<400> 64
His Ser Val Arg Ser Phe Arg His Arg Asn Ala Asn Thr Phe Leu Thr
1 5 10 15

Tyr Ser

<210> 65
<211> 18
<212> PRT
<213> Mastrevirus

<400> 65
His Ser Val Arg Ser Phe Arg His Arg Asn Ala Asn Thr Phe Leu Thr
1 5 10 15

Tyr Ser

<210> 66
<211> 17
<212> PRT
<213> Mastrevirus

<400> 66
Pro Ser Arg Arg Phe Lys His Arg Asn Val Asn Thr Phe Leu Thr Tyr
1 5 10 15

Ser

<210> 67
<211> 16
<212> PRT
<213> Mastrevirus

<400> 67
Thr Lys Ser Phe Arg Leu Gln Thr Lys Tyr Val Phe Leu Thr Tyr Pro
1 5 10 15

<210> 68
<211> 15
<212> PRT
<213> Mastrevirus

<400> 68
Pro Arg Phe Arg Val Tyr Ser Lys Tyr Leu Phe Leu Thr Tyr Pro
1 5 10 15

<210> 69
<211> 15
<212> PRT
<213> Mastrevirus

<400> 69
Pro Arg Phe Arg Val Tyr Ser Lys Tyr Leu Phe Leu Thr Tyr Pro
1 5 10 15

<210> 70
<211> 20
<212> PRT
<213> Curtovirus

<400> 70
Met Pro Pro Thr Lys Arg Phe Arg Ile Gln Ala Lys Asn Ile Phe Leu
1 5 10 15

Thr Tyr Pro Gln
20

<210> 71
<211> 21
<212> PRT

<213> Curtovirus

<400> 71
Met Pro Arg Thr Pro Lys Arg Phe Arg Ile Gln Ala Lys Asn Ile Phe
1 5 10 15

Leu Thr Tyr Pro Gln
20

<210> 72
<211> 16
<212> PRT
<213> Curtovirus

<400> 72
Met Pro Phe Tyr Lys Lys Ala Lys Asn Phe Phe Leu Thr Tyr Pro Gln
1 5 10 15

<210> 73
<211> 20
<212> PRT
<213> Curtovirus

<400> 73
Met Pro Arg Ser Pro Ser Phe Gln Ile Lys Ala Lys Asn Ile Phe Leu
1 5 10 15

Thr Tyr Pro Arg
20

<210> 74
<211> 21
<212> PRT
<213> Curtovirus

<400> 74
Met Pro Arg Gln Pro Asn Ser Phe Arg Ile Gln Ala Arg Asn Ile Phe
1 5 10 15

Leu Thr Tyr Pro Gln
20

<210> 75
<211> 21
<212> PRT
<213> Curtovirus

<400> 75
Met Pro Ser Asn Pro Lys Arg Phe Gln Ile Ala Ala Lys Asn Tyr Phe
1 5 10 15

Leu Thr Tyr Pro Asn
20

<210> 76
<211> 19
<212> PRT

<213> Begomovirus

<400> 76
Met Pro Pro Pro Lys Arg Phe Gln Ile Asn Ser Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 77
<211> 19
<212> PRT
<213> Begomovirus

<400> 77
Met Ala Pro Pro Lys Arg Phe Gln Ile Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 78
<211> 19
<212> PRT
<213> Begomovirus

<400> 78
Met Pro Arg Ala Gly Arg Phe Gln Ile Asn Ala Lys Asn Tyr Phe Ile
1 5 10 15

Thr Tyr Pro

<210> 79
<211> 19
<212> PRT
<213> Begomovirus

<400> 79
Met Pro Arg Ala Gly Arg Phe Gln Ile Asn Ala Lys Asn Tyr Phe Val
1 5 10 15

Thr Tyr Pro

<210> 80
<211> 19
<212> PRT
<213> Begomovirus

<400> 80
Met Pro Arg Ala Gly Arg Phe Gln Ile Asn Ala Lys Asn Tyr Phe Ile
1 5 10 15

Thr Tyr Pro

<210> 81
<211> 19
<212> PRT
<213> Begomovirus

<400> 81
Met Ser Pro Pro Lys Arg Phe Gln Ile Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 82
<211> 19
<212> PRT
<213> Begomovirus

<400> 82
Met Ala Pro Pro Lys Gln Phe Gln Ile Tyr Ala Lys Asn Tyr Phe Ile
1 5 10 15

Thr Tyr Pro

<210> 83
<211> 18
<212> PRT
<213> Begomovirus

<400> 83
Met Pro Pro Lys Arg Phe Leu Ile Asn Ser Lys Asn Tyr Phe Leu Thr
1 5 10 15

Tyr Pro

<210> 84
<211> 20
<212> PRT
<213> Begomovirus

<400> 84
Met Pro Ser His Pro Lys Arg Phe Gln Ile Asn Ala Lys Asn Tyr Phe
1 5 10 15

Leu Thr Tyr Pro
20

<210> 85
<211> 19
<212> PRT
<213> Begomovirus

<400> 85
Met Ala Pro Pro Lys Arg Phe Gln Ile Asn Cys Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 86
<211> 19
<212> PRT
<213> Begomovirus

<400> 86
Met Ala Gln Pro Lys Arg Phe Gln Ile Asn Ala Lys His Tyr Phe Leu
1 5 10 15

Ser Phe Pro

<210> 87
<211> 19
<212> PRT
<213> Begomovirus

<400> 87
Met Ala Gln Pro Lys Arg Phe Gln Ile Asn Ala Lys His Tyr Phe Leu
1 5 10 15

Thr Phe Pro

<210> 88
<211> 19
<212> PRT
<213> Begomovirus

<400> 88
Met Pro Arg Ala Gly Arg Phe Gln Ile Asn Ala Lys Asn Tyr Phe Ile
1 5 10 15

Thr Tyr Pro

<210> 89
<211> 19
<212> PRT
<213> Begomovirus

<400> 89
Met Pro Arg Asn Asn Ser Phe Cys Ile Asn Ala Lys Asn Ile Phe Leu
1 5 10 15

Thr Phe Pro

<210> 90
<211> 19
<212> PRT
<213> Begomovirus

<400> 90
Met Pro Arg Asn Asn Ser Phe Cys Ile Asn Ala Lys Asn Ile Phe Leu

1 5 10 15

Thr Phe Pro

<210> 91
<211> 19
<212> PRT
<213> Begomovirus

<400> 91
Met Pro Arg Leu Asn Ser Phe Cys Val Asn Ala Lys Asn Ile Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 92
<211> 19
<212> PRT
<213> Begomovirus

<400> 92
Met Ala Ala Pro Asn Arg Phe Lys Leu Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 93
<211> 19
<212> PRT
<213> Begomovirus

<400> 93
Met Pro Arg Lys Gly Ser Phe Ser Val Lys Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 94
<211> 19
<212> PRT
<213> Begomovirus

<400> 94
Met Pro Pro Pro Lys Arg Phe Ser Val Asn Ala Lys Asn Phe Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 95
<211> 19
<212> PRT
<213> Begomovirus

<400> 95
Met Pro Arg Lys Gly Ser Phe Ser Ile Lys Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 96
<211> 19
<212> PRT
<213> Begomovirus

<400> 96
Met Pro Arg Lys Gly Ser Phe Ser Ile Lys Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 97
<211> 19
<212> PRT
<213> Begomovirus

<400> 97
Met Pro Arg Lys Gly Tyr Phe Ser Val Lys Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 98
<211> 19
<212> PRT
<213> Begomovirus

<400> 98
Met Pro Arg Ser Gly Arg Phe Ser Ile Lys Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro

<210> 99
<211> 11
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (6)
<223> x can be 3-30 nucleotides

<400> 99
ggagaxggag a 11

<210> 100

<211> 11
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (6)
<223> x can be 3-30 nucleotides

<400> 100
ggtgtxggtg t 11

<210> 101
<211> 19
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (10)
<223> x can be 3-30 nucleotides

<400> 101
ggtgcgcccx gggcgcacc 19

<210> 102
<211> 18
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (9)
<223> x can be 3-30 nucleotides

<400> 102
gcgccttcxg aaggcgcg 18

<210> 103
<211> 17
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (9)
<223> x can be 3-30 nucleotides

<400> 103
ggtttgcgxc gcaaacc 17

<210> 104
<211> 15
<212> DNA
<213> Geminivirus

<220>
<221> unsure

<222> (8)
<223> x can be 3-30 nucleotides

<400> 104
gtgagtgxca ctcac 15

<210> 105
<211> 23
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (13)
<223> x can be 3-30 nucleotides

<400> 105
ggaggtgcgt ccxcctccac ggg 23

<210> 106
<211> 11
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (6)
<223> x can be 3-30 nucleotides

<400> 106
ggagtxggag t 11

<210> 107
<211> 11
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (6)
<223> x can be 3-30 nucleotides

<400> 107
ggtacxggta c 11

<210> 108
<211> 11
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (6)
<223> x can be 3-30 nucleotides

<400> 108
ggtacxggta c 11

<210> 109
<211> 11
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (6)
<223> x can be 3-30 nucleotides

<400> 109
ggggaxgggg a 11

<210> 110
<211> 11
<212> DNA
<213> Geminivirus

<220>
<221> unsure
<222> (6)
<223> x can be 3-30 nucleotides

<400> 110
gggggxgggg g 11

<210> 111
<211> 13
<212> DNA
<213> Geminivirus

<400> 111
ggtgtctgga gtc 13

<210> 112
<211> 11
<212> PRT
<213> Geminivrus

<400> 112
Ala Asn Val Arg Phe Arg Arg Pro Pro Ala Met
1 5 10

<210> 113
<211> 11
<212> PRT
<213> Geminivrus

<400> 113
Ala Asp Ile Arg Phe Arg Arg Pro Pro Ala Met
1 5 10

<210> 114
<211> 11
<212> PRT
<213> Geminivrus

<400> 114
Ala Asn Val Arg Phe Arg Arg Pro Pro Ala Met
1 5 10

<210> 115
<211> 11
<212> PRT
<213> Geminivrus

<400> 115
Ala Asn Ile Arg Phe Arg Arg Pro Pro Ala Met
1 5 10

<210> 116
<211> 11
<212> PRT
<213> Geminivrus

<400> 116
Ala Asn Val Arg Phe Arg Arg Pro Pro Ala Met
1 5 10

<210> 117
<211> 11
<212> PRT
<213> Geminivrus

<400> 117
Ala Asn Val Arg Phe Arg Arg Pro Pro Ala Met
1 5 10

<210> 118
<211> 11
<212> PRT
<213> Geminivrus

<400> 118
Ala Asp Val Arg Phe Arg Arg Pro Pro Ala Met
1 5 10

<210> 119
<211> 11
<212> PRT
<213> Geminivrus

<400> 119
Ala Asn Val Arg Phe Arg Arg Pro Pro Ala Met
1 5 10

<210> 120
<211> 11
<212> PRT
<213> Geminivrus

<400> 120
Ala Asp Val Arg Phe Arg Arg Pro Pro Ala Met

1 5 10

<210> 121
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 121
ggtgtctgga gtc 13

<210> 122
<211> 13
<212> DNA
<213> Geminivrus

<400> 122
ggcgtctggc gtc 13

<210> 123
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligomucleotide

<400> 123
gggtctggag tc 12

<210> 124
<211> 12
<212> DNA
<213> Geminivrus

<400> 124
ggtgtctggg tc 12

<210> 125
<211> 13
<212> DNA
<213> Geminivrus

<400> 125
ggcgtctgga gtc 13

<210> 126
<211> 4
<212> PRT
<213> Geminivrus

<400> 126

Phe Arg Val Gln
1

<210> 127
<211> 4
<212> PRT
<213> Geminivrus

<400> 127
Phe Arg Val Asn
1

<210> 128
<211> 4
<212> PRT
<213> Geminivrus

<400> 128
Phe Arg Ile Asn
1

<210> 129
<211> 4
<212> PRT
<213> Geminivrus

<400> 129
Phe Arg Ile Gln
1

<210> 130
<211> 4
<212> PRT
<213> Geminivrus

<400> 130
Phe Arg Leu Gln
1

<210> 131
<211> 4
<212> PRT
<213> Geminivrus

<400> 131
Phe Lys Val Gln
1

<210> 132
<211> 4
<212> PRT
<213> Geminivrus

<400> 132
Phe Lys Ile Tyr
1

<210> 133
<211> 4
<212> PRT
<213> Geminivrus

<400> 133
Phe Lys Ile Asn
1

<210> 134
<211> 4
<212> PRT
<213> Geminivrus

<400> 134
Phe Arg Leu Ala
1

<210> 135
<211> 4
<212> PRT
<213> Geminivrus

<400> 135
Phe Arg Leu Asn
1

<210> 136
<211> 4
<212> PRT
<213> Geminivrus

<400> 136
Leu Lys Thr Asn
1

<210> 137
<211> 4
<212> PRT
<213> Geminivrus

<400> 137
Phe Ala Ile Asn
1

<210> 138
<211> 4
<212> PRT
<213> Geminivrus

<400> 138
Phe Arg Leu Thr
1

<210> 139

<211> 4
<212> PRT
<213> Geminivrus

<400> 139
Phe Asn Ile Asn
1

<210> 140
<211> 4
<212> PRT
<213> Geminivrus

<400> 140
Phe Arg Val Asn
1

<210> 141
<211> 4
<212> PRT
<213> Geminivrus

<400> 141
Phe Ser Ile Asn
1

<210> 142
<211> 4
<212> PRT
<213> Geminivrus

<400> 142
Phe Lys Ile Tyr
1

<210> 143
<211> 4
<212> PRT
<213> Geminivrus

<400> 143
Phe Leu Ile Asn
1

<210> 144
<211> 4
<212> PRT
<213> Geminivrus

<400> 144
Phe Gln Ile Asn
1

<210> 145
<211> 4
<212> PRT

<213> Geminivrus

<400> 145
Phe Gln Ile Tyr
1

<210> 146
<211> 4
<212> PRT
<213> Geminivrus

<400> 146
Phe Cys Ile Asn
1

<210> 147
<211> 4
<212> PRT
<213> Geminivrus

<400> 147
Phe Cys Val Asn
1

<210> 148
<211> 4
<212> PRT
<213> Geminivrus

<400> 148
Phe Lys Leu Asn
1

<210> 149
<211> 4
<212> PRT
<213> Geminivrus

<400> 149
Phe Ser Val Lys
1

<210> 150
<211> 4
<212> PRT
<213> Geminivrus

<400> 150
Phe Ser Val Asn
1

<210> 151
<211> 4
<212> PRT
<213> Geminivrus

<400> 151
Phe Ser Ile Lys
1

<210> 152
<211> 4
<212> PRT
<213> Geminivrus

<400> 152
Phe Tyr Lys Lys
1

<210> 153
<211> 4
<212> PRT
<213> Geminivrus

<400> 153
Phe Gln Ile Lys
1

<210> 154
<211> 4
<212> PRT
<213> Geminivrus

<400> 154
Phe Gln Ile Ala
1

<210> 155
<211> 7
<212> PRT
<213> Geminivrus

<400> 155
Phe Arg Leu Gln Thr Lys Tyr
1 5

<210> 156
<211> 7
<212> PRT
<213> Geminivrus

<400> 156
Phe Arg Val Tyr Ser Lys Tyr
1 5

<210> 157
<211> 6
<212> PRT
<213> Geminivrus

<400> 157
His Arg Asn Ala Asn Thr

1 5

<210> 158
<211> 4
<212> DNA
<213> Geminivrus

<400> 158
tata 4

<210> 159
<211> 26
<212> DNA
<213> Geminivrus

<400> 159
ggcgtctggc gtcggtgtct ggagtc 26

<210> 160
<211> 14
<212> DNA
<213> Geminivrus

<400> 160
ggtgtctggc gtct 14

<210> 161
<211> 14
<212> DNA
<213> Geminivrus

<400> 161
ggagtctggc gtct 14

<210> 162
<211> 14
<212> DNA
<213> Geminivrus

<400> 162
ggagtctggt gtct 14

<210> 163
<211> 14
<212> DNA
<213> Geminivrus

<400> 163
ggcgtctgga gtct 14

<210> 164
<211> 6
<212> DNA
<213> Geminivrus

<400> 164
ggtgtc 6

<210> 165
<211> 13
<212> DNA
<213> Geminivrus

<400> 165
ggtgtctggt gtc 13

<210> 166
<211> 27
<212> DNA
<213> Geminivrus

<400> 166
ggtgtctggt gtctggtgtc tggtgtc 27

<210> 167
<211> 7
<212> PRT
<213> Geminivrus

<400> 167
Phe Leu Thr Tyr Pro Lys Cys
1 5

<210> 168
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 168
ggtgtctggc gtc 13

<210> 169
<211> 13
<212> DNA
<213> Geminivrus

<400> 169
ggggtctggc gtc 13

<210> 170
<211> 12
<212> DNA
<213> Geminivrus

<400> 170
gggtctggtg tc 12

<210> 171
<211> 13
<212> DNA
<213> Geminivrus

<400> 171
ggcgtctggg gtc 13

<210> 172
<211> 18
<212> DNA
<213> Geminivrus

<400> 172
ggtgtctttt ttggagtc 18

<210> 173
<211> 18
<212> DNA
<213> Geminivrus

<400> 173
ggcgtctttt ttggcgtc 18

<210> 174
<211> 12
<212> DNA
<213> Geminivrus

<400> 174
ggtgtcggag tc 12

<210> 175
<211> 12
<212> DNA
<213> Geminivrus

<400> 175
ggcgtcggcg tc 12

<210> 176
<211> 6
<212> DNA
<213> Geminivrus

<400> 176
ggcgtc 6

<210> 177
<211> 27
<212> DNA
<213> Geminivrus

<400> 177
ggtgtctgga gtctggtgtc tggagtc 27

<210> 178
<211> 27
<212> DNA
<213> Geminivrus

<400> 178
ggcgtctggc gtctggcgtc tggcgtc 27

<210> 179
<211> 13
<212> DNA
<213> Geminivrus

<400> 179
ggagtctgga gtc 13

<210> 180
<211> 16
<212> DNA
<213> Geminivrus

<400> 180
ggcgtctggc gtctac 16

<210> 181
<211> 16
<212> DNA
<213> Geminivrus

<400> 181
ggtgtctgga gtctac 16

<210> 182
<211> 13
<212> DNA
<213> Geminivrus

<400> 182
ggggtctgga gtc 13

<210> 183
<211> 361
<212> PRT
<213> Geminivrus

<400> 183
Met Ala Ser Pro Arg Arg Phe Arg Val Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro Lys Cys Ser Leu Thr Lys Glu Glu Ala Leu Ser Gln Leu
20 25 30

Gln Thr Leu Glu Thr Pro Thr Lys Lys Lys Phe Ile Lys Ile Cys Arg
35 40 45

Glu Leu His Glu Asp Gly Ser Pro His Ile His Val Leu Ile Gln Phe
50 55 60

```
Glu Gly Lys Phe Gln Cys Lys Asn Asn Arg Phe Phe Asp Leu Val Ser
 65                  70                  75                  80

Pro Ser Arg Ser Ala His Phe His Pro Asn Ile Gln Gly Ala Lys Ser
                 85                  90                  95

Ala Ser Asp Val Lys Asn Tyr Ile Ala Lys Asp Gly Asp Val Leu Glu
            100                 105                 110

Trp Gly Val Phe Gln Ile Asp Gly Arg Ser Ala Arg Gly Gly Gln Gln
        115                 120                 125

Thr Ala Asn Asp Ala Tyr Ala Gln Ala Ile Asn Thr Gly Asn Lys Asp
    130                 135                 140

Asp Ala Leu Arg Val Leu Lys Glu Leu Ala Pro Lys Asp Tyr Val Leu
145                 150                 155                 160

Gln Phe His Asn Leu Asn Thr Asn Leu Asp Arg Ile Phe Gln Pro Pro
                165                 170                 175

Ser Glu Val Tyr Val Ser Pro Phe Ser Ile Ser Ser Phe Asp Arg Val
            180                 185                 190

Pro Ala Asp Leu Val Asp Trp Val Ser Ser Asn Val Val Cys Ala Ala
        195                 200                 205

Ala Arg Pro Phe Arg Pro Ile Ser Ile Val Ile Glu Gly Asp Ser Arg
    210                 215                 220

Thr Gly Lys Thr Met Trp Ala Arg Cys Leu Gly Pro His Asn Tyr Leu
225                 230                 235                 240

Cys Gly His Leu Asp Leu Ser Pro Lys Val Tyr Ser Asn Asp Ala Trp
                245                 250                 255

Tyr Asn Val Ile Asp Asp Val Asp Pro His Tyr Leu Lys His Phe Lys
            260                 265                 270

Glu Phe Met Gly Ala Gln Arg Asp Cys Gln Ser Asn Thr Lys Tyr Gly
        275                 280                 285

Lys Pro Val Met Ile Lys Gly Gly Ile Pro Thr Ile Phe Leu Cys Asn
    290                 295                 300

Lys Gly Pro Asn Ser Ser Tyr Lys Glu Tyr Leu Asp Glu Glu Lys Asn
305                 310                 315                 320

Ala Ala Leu Lys Gln Trp Ala Ile Lys Asn Ala Val Phe Ile Thr Leu
                325                 330                 335

Glu Glu Pro Leu Tyr Ser Gly Arg Glu Asn Ile Ala Leu Pro Glu Glu
            340                 345                 350

Glu Glu Glu His Ser Gln Glu Ala Ser
        355                 360
```

<210> 184
<211> 52
<212> PRT
<213> Geminivrus

<400> 184
Met Ala Ser Pro Arg Arg Phe Arg Val Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro Lys Cys Ser Leu Thr Lys Glu Glu Ala Leu Ser Gln Leu
20 25 30

Gln Thr Leu Glu Thr Pro Thr Lys Lys Lys Phe Ile Lys Ile Cys Arg
35 40 45

Glu Leu His Glu
50

<210> 185
<211> 114
<212> PRT
<213> Geminivrus

<400> 185
Met Ala Ser Pro Arg Arg Phe Arg Val Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro Lys Cys Ser Leu Thr Lys Glu Glu Ala Leu Ser Gln Leu
20 25 30

Gln Thr Leu Glu Thr Pro Thr Lys Lys Lys Phe Ile Lys Ile Cys Arg
35 40 45

Glu Leu His Glu Asp Gly Ser Pro His Ile His Val Leu Ile Gln Phe
50 55 60

Glu Gly Lys Phe Gln Cys Lys Asn Asn Arg Phe Phe Asp Leu Val Ser
65 70 75 80

Pro Ser Arg Ser Ala His Phe His Pro Asn Ile Gln Gly Ala Lys Ser
85 90 95

Ala Ser Asp Val Lys Asn Tyr Ile Ala Lys Asp Gly Asp Val Leu Glu
100 105 110

Trp Gly

<210> 186
<211> 160
<212> PRT
<213> Geminivrus

<400> 186
Met Ala Ser Pro Arg Arg Phe Arg Val Asn Ala Lys Asn Tyr Phe Leu
1 5 10 15

Thr Tyr Pro Lys Cys Ser Leu Thr Lys Glu Glu Ala Leu Ser Gln Leu
20 25 30

Gln Thr Leu Glu Thr Pro Thr Lys Lys Lys Phe Ile Lys Ile Cys Arg
35 40 45

Glu Leu His Glu Asp Gly Ser Pro His Ile His Val Leu Ile Gln Phe
50 55 60

Glu Gly Lys Phe Gln Cys Lys Asn Asn Arg Phe Phe Asp Leu Val Ser
65 70 75 80

Pro Ser Arg Ser Ala His Phe His Pro Asn Ile Gln Gly Ala Lys Ser
85 90 95

Ala Ser Asp Val Lys Asn Tyr Ile Ala Lys Asp Gly Asp Val Leu Glu
100 105 110

Trp Gly Val Phe Gln Ile Asp Gly Arg Ser Ala Arg Gly Gly Gln Gln
115 120 125

Thr Ala Asn Asp Ala Tyr Ala Gln Ala Ile Asn Thr Gly Asn Lys Asp
130 135 140

Asp Ala Leu Arg Val Leu Lys Glu Leu Ala Pro Lys Asp Tyr Val Leu
145 150 155 160

<210> 187
<211> 9
<212> DNA
<213> Geminivrus

<400> 187
taatattac 9

<210> 188
<211> 5
<212> DNA
<213> Geminivrus

<400> 188
ggtac 5

<210> 189
<211> 4
<212> PRT
<213> Geminivrus

<400> 189
Phe Leu Thr Tyr

<210> 190
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 190
ggcgtctttt ttggcgtc 18

<210> 191
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 191
ggtgtctggt gtctggtgtc tggtgtc 27

<210> 192
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonulceoide

<400> 192
ggcgtctggc gtctggcgtc tggcgtc 27

<210> 193
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotides

<400> 193
ggagtctgga gtc 13

<210> 194
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotides

<400> 194
ggtgtctggt gtc 13

<210> 195
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotides

<400> 195
ggagtctgga gtc 13

<210> 196
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotides

<400> 196
ggtgtctgga gtc 13

<210> 197
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotides

<400> 197
ggcgtctggc gtc 13

<210> 198
<211> 6
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotides

<400> 198
ggtgtc 6

<210> 199
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotides

<400> 199
ggtgtctgga gtctggtgtc tggagtc 27

<210> 200
<211> 27
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic oligonucleotides

<400> 200
ggcgtctggc gtctggcgtc tggcgtc 27

<210> 201
<211> 13
<212> DNA
<213> Geminivrus

<400> 201
ggtgtctggt gtc 13

<210> 202
<211> 13
<212> DNA
<213> Geminivrus

<400> 202
ggagtctgga gtc 13

<210> 203
<211> 13
<212> DNA
<213> Geminivrus

<400> 203
ggtgtctggt gtc 13

<210> 204
<211> 6
<212> DNA
<213> Geminivrus

<400> 204
ggtgtc 6

<210> 205
<211> 13
<212> DNA
<213> Geminivrus

<400> 205
ggtgtctgga gtc 13

<210> 206
<211> 5
<212> DNA
<213> Geminivrus

<400> 206
ggaga 5

<210> 207

<211> 5
<212> DNA
<213> Geminivrus

<400> 207
ggtga 5

<210> 208
<211> 5
<212> DNA
<213> Geminivrus

<400> 208
ggtgt 5

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 July 2000 (27.07.2000)

PCT

(10) International Publication Number
WO 00/43494 A3

(51) International Patent Classification[7]: C12N 5/04, 15/33, 15/63, 15/82, 15/83, 15/84, 15/90, A01H 5/00, 5/10

(21) International Application Number: PCT/US00/01849

(22) International Filing Date: 27 January 2000 (27.01.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/117,285 26 January 1999 (26.01.1999) US

(71) Applicant *(for all designated States except US)*: **THE SCRIPPS RESEARCH INSTITUTE** [US/US]; Office of Patent Counsel, 10550 North Torrey Pines Road, Mail TPC-8, La Jolla, CA 92037 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FAUQUET, Claude** [FR/US]; 308 South Spoede Road, St. Louis, MO 63141 (US). **CHATTERJI, Anju** [IN/US]; 613C Westgate Avenue, University City, MO 63130 (US).

(74) Agents: **ESMOND, Robert, W.** et al.; Sterne, Kessler, Goldstein & Fox P.L.L.C., Suite 600, 1100 New York Avenue, N.W., Washington, DC 20005-3934 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *With international search report.*

(88) Date of publication of the international search report:
30 November 2000

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 00/43494 A3

(54) Title: CONTROL OF VIRUS INFECTION USING REPLICATION ASSOCIATED PROTEINS, COMPOSITIONS AND METHODS OF USE

(57) Abstract: The invention is generally directed to methods and compositions for controlling infection of plants by geminiviruses using replication associated proteins. In particular, the invention relates to iteron and Rep DNA and polypeptide sequences as well as transgenic plants expressing such sequences.

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/01849

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 320.1, 410, 419, 468-470; 536/23.72; 800.278, 279, 295, 298, 301, 308, 310-315, 317.3, 320, 320.1-320.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

West, Agricola, Caplus, Biosis

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A, T | PAXIMADIS et al. Characterization Of Tobacco Geminivirus In The Old And New World. Arch. Virol. 1999. Vol. 144. pages 703-717, especially pages 710-713. | 1, 3, 8-12, 14-18, 30-38, 43, 48, 491-3 |
| Y | HORSCH et al. A Simple And General Method For Transferring Genes Into Plants. Science. 08 March 1985. Vol. 227. pages 1229-1231, see entire document. | 8, 12 |
| Y | SANFORD et al. Optimizing the Biolistic Process For Different Biological Applications. Meth. Enzymol. 1993. Vol. 217. pages 483-509, see entire document. | 8, 10, 11 |

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

| | | | |
|---|---|---|---|
| * | Special categories of cited documents: | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "A" | document defining the general state of the art which is not considered to be of particular relevance | "X" | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "E" | earlier document published on or after the international filing date | "Y" | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "L" | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "&" | document member of the same patent family |
| "O" | document referring to an oral disclosure, use, exhibition or other means | | |
| "P" | document published prior to the international filing date but later than the priority date claimed | | |

| | |
|---|---|
| Date of the actual completion of the international search<br>22 MAY 2000 | Date of mailing of the international search report<br>13 JUL 2000 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230 | Authorized officer<br>ASHWIN MEHTA<br>Telephone No. (703) 308-0196 |

Form PCT/ISA/210 (second sheet) (July 1998) *

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/01849

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | ARGUELLO-ASTORGA et al. Geminivirus Replication Origins Have A Group-Specific Organization Of Iterative Elements: A Model For Replication. Virol. 1994. Vol. 203. pages 90-100, especially pages 94-98. | 1, 3, 8-12, 14-18, 30-38, 43, 48, 49 |
| Y | ARGUELLO-ASTORGA et al. Experimental And Theoretical Definition Of Geminivirus Origin Of Replication. Plant Mol. Biol. 1994. Vol. 26. pages 553-556, especially pages 554-555. | 1, 3, 8-12, 14-18, 43, 48, 49 |

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/01849

**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-3, 5, 8-12, 14-20, 30-38, 43, 48, 49 to the extent of Species A.

**Remark on Protest**

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/01849

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (7):

C12N 5/04, 15/33, 15/63, 15/82, 15/83, 15/84, 15/90; A01H 5/00, 5/10

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/69.1, 320.1, 410, 419, 468-470; 536/23.72; 800.278, 279, 295, 298, 301, 308, 310-315, 317.3, 320, 320.1-320.3

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s)1-38,40, 43, 48, 49, 57-61, drawn to a first method, for producing resistance in a plant to a geminivirus, comprising introducing a geminivirus Rep-iteron antagonist wherein said antagonist is selected from the group consisting of a nucleotide sequence of a geminivirus iteron capable of binding to a Rep protein, and a defective Rep protein, and a first product, comprising nucleic acid sequences encoding a gemninivirus antagonist.
Group II, claim(s) 39, 41, 42, 46, 47, and 50, drawn to a second product, polypeptides encoding a defective Rep protein.
Group III, claim(s) 44 and 45, drawn to a second method, for providing resistance to infection by geminiviruses comprising transforming plant cells with a gene encoding a defective Rep protein.
Group IV, claim(s)51-56, drawn to a third method, comprising producing at least partial resisitance to a virus in a plant or animal cell.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be searched, the appropriate additional search fees must be paid. The species are as follows:

In regard to Group I: Species A- a method for producing resistance to geminivirus comprising introducing a nucleotide sequence encoding a geminivirus iteron; Species B- a method for producing resistance to geminivirus comprising introducing a nucleotide sequence encoding a defective Rep protein; Species C- a Rep iteron antagonist comprising a polypeptide.
The claims are deemed to correspond to the species listed above in the following manner:

Species A- claims 1-3, 5, 8-12, 14-20, 30-38, 43, 48, and 49; Species B- claims 1, 3, 4, 6, 7, 9-18, 21-29, 37, 38, 40, 43, 48, 57-61
Species C- claims 59 and 60.

The following claims are generic: claims 1, 3, 9-18, 43, 48.

The inventions listed as Groups I-IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the method and nucleotide sequences of of Group I are not shared with the proteins of Group II. The iterons of Group I are not shared with the method of Group III. The proteins of Group II are not shared with the other Groups. The method of Group III is not shared with the polypeptides of Group II nor the animal cells of Group IV. The method comprising virus resistance in animal cells of Group IV is not shared with the other groups.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: The species encode different nucleic acid sequences. The iteron-encoding sequences of Species
A are not shared with defective Rep-encoding sequences of Species B. Species C is drawn to amino acid sequences that are not shared with Species A nor B.